Faculty of Arts & Philosophy

Troy Devon Thomas

Response Styles and the Quality of Survey Data

Evidence from Guyana

Dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Comparative Science of Cultures

2014

Promoter	Prof. Dr. Koenraad Stroeken
Co-promoter	Dr. Koen Abts [University of Leuven]

Decaan	Prof. dr. Marc Boone
Rector	Prof. dr. Anne De Paepe

Funding for this dissertation was provided by the Flemish Interuniversity Council. Grant No. ZEIN2008PR357

To my wife who has made significant sacrifices to facilitate my studies,
my unborn son whom I hope will improve on even the things
that I believe that I've done optimally, and my
parents for giving education priority.

ACKNOWLEDGMENT

In the process of completing this dissertation, I came into contact with both organisations and individuals who have in one way or another assisted me. I wish to say thank you.

My PhD studies were made possible by a collaborative project between the University of Ghent and the University of Guyana and it was funded by the Flemish Interuniversity Council. I am grateful for the support of these organisations and that of the individuals within them who have assisted me. I wish to thank Dr. Michael Scott and Oneil Greaves of the University of Guyana for their support from the Guyana end.

I say thank you to my supervisors who have managed the project and dedicated a lot of time and effort to providing guidance. Thank you to Prof. dr. Patrick Vander Weyden who was instrumental in starting the project, recruited me for PhD studies and was my supervisor while I was attached to the Department of Political Science. Thanks to Prof. dr. Koenraad Stroeken who was my supervisor during the latter part of the project when I became attached to the Department of Comparative Science of Cultures. Thanks to Dr. Koen Abts who co-supervised my work. Thanks for putting considerable time and effort into commenting on drafts and for suggesting ideas. The quality of my work improved as a result of your input. A good supervisory team was important to my completing this dissertation and having thanked you individually, I wish to thank you collectively.

At the start of my PhD studies, I had not decided on a topic. My first contact with response styles came through Prof. dr. Jaak Billiet while I was a student on one of his courses. The topic piqued my interest, but my decision to focus on it came after Prof. dr. Yves Van Vaerenbergh who was also a PhD student at the time, approached me with the proposal to collaborate. I wish to say a heart-felt thank you to Yves and to Prof. Billiet for introducing me the topic which has become my core area of research. On a related note, I also wish to thank Prof. dr. Bert Weijters whom I have not met in person up to the time of writing this acknowledgement. However, Prof. Weijters never failed to respond to my questions about measuring response styles. This was particularly helpful in the early stages of my PhD. I also wish to thank Prof. dr. Geert Molenbergh for answering my many questions about the sampling methodology.

Several of my friends and colleagues were also supportive of my work. I thank Prof. dr. Nancy Fraser (Seneca College, Toronto) for assisting me with accessing resources that I needed especially in the early stages when I was working in Guyana. I say thank you to Jorg Vereecke who has been involved in my project form the start and who continued to provide support throughout. Thanks to Lenandlar Singh and Malcolm Williams of the University of Guyana for reading and commenting on drafts of the early articles. Thanks to Max Baird (University of Guyana) for providing technical support throughout my studies and for lending me his equipment whenever I needed extra resources. I also wish to thank Roland Daynauth and Dwayne Renville (University of Guyana) for their assistance during my period of study.

To my family, I wish to say thanks for your support. The decisions that were made by my Mom and Dad while I was a child have impacted greatly on my life. That I am able to take on PhD studies at all is ultimately a result of your difficult and wise choices for your children. I hope that you find reward in what we are able to accomplish in adulthood. To my brother, Erwin Thomas who took charge of my affairs whenever I needed to be away, I wish to say that you have been a "big help". With you in charge my mind was always at ease and I could concentrate on what I was doing. To my wife, Nina, I wish to indicate that you have in many ways enabled me to focus on my work. You were quite understanding and supportive and have been more than I could reasonably expect.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Introduction

Surveys are an important source of research data in many fields. They are particularly indispensable to social research which contributes not only to theory, but also to practice and policy formulation. It is therefore important to control errors that arise in survey research. In general, there are four sources of errors in surveys (Schwarz, Groves, & Schuman, 1998):

Coverage Error occurs when the sample frame is not a true representation of the population.

Sampling Error occurs when it is not possible to conduct a census leading to the necessity of studying a sample of the population.

Nonresponse Error results from identified respondents are not included due to refusal, non-contact or other issues that result in non-participation.

Measurement Error occurs when there is a difference between the respondent's true score on a construct and his/her observed score. This results from a mismatch between the response provided and the respondent's true opinion.

This dissertation focuses on measurement error and specifically on the component of systematic measurement error which is referred to as response styles (RSs).

RSs are the respondents' systematic tendencies to respond in certain ways to rating scale items regardless of the content of the items (Paulhus, 1991). The most popular RSs are acquiescence RS (ARS: tendency to agree) and extreme RS (ERS: tendency to use the scale endpoints). These two RSs have received most attention from researchers and consequently, most of what is known about response styles is relevant to ARS and ERS. However, disacquiescence RS (DARS: tendency to disagree) and midpoint RS (MRS: tendency to use the scale midpoint) are still well-recognised. Although, RSs emerge with the use of rating scales, rating scales remain a prominent feature of survey questionnaires (Moors, 2010). Rating scales are useful and often preferred because they are easy to use and are easily combined into batteries (Krosnick, 1999). The popularity of such scales in survey research underscores the importance of understanding, measuring and controlling RSs (Moors, 2003).

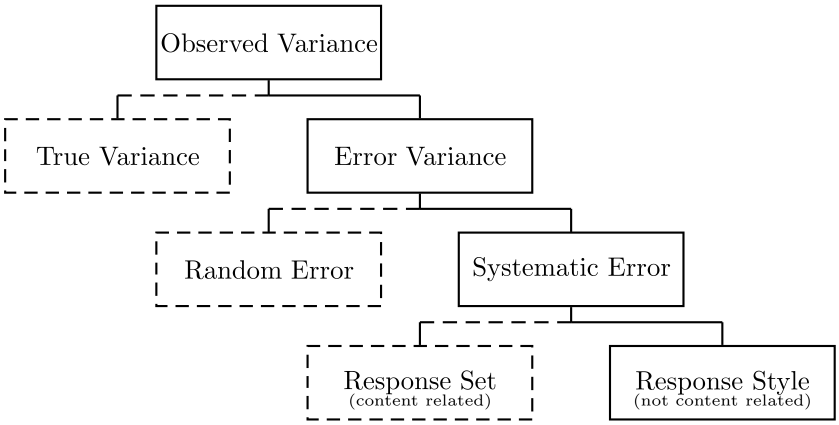

Figure 1.1: Decomposition of Observed Variance[1]

Response styles are known to affect the variance of rating scale items. At the data analysis stage, the observed variance of a item consists of the true variance and error variance (Smith, 2011). The true variance is the component that the item is intended to measure, but the ideal state in which the variance of an item equals its true variance is not achieved due to errors (Figure 1.1). Error variance is further decomposed into a random and a systematic component.

Variations caused by random error are due to chance and are not very problematic. Random error can be dealt with by, for example, using multi-item scales and post-hoc assessment of reliability (Churchill, 1979; Cronbach, 1951). Systematic error on the other hand, imply that the resulting variations are predictable and this poses serious problems. Two examples of systematic errors are response set and RS. Response set is related to the content of the items and it conveys a sense of impression management. Socially desirable responding is an example of a response set. In contrast to response set, an RS is unrelated

[1]Taken from Weijters (2006)

to the content of the items and is reflective of cognitive rather than social processes (Ayidiya & McClendon, 1990). RSs result in systematic mismatches between the registered responses and the respondents' true opinions. They bias the results of survey research and routinely lead to incorrect conclusions (Baumgartner & Steenkamp, 2001; Moors, 2012). RSs therefore affect the quality of survey data which we define as the accuracy of research results.

In general, data quality is affected by culture. However, data quality research is done almost entirely in Western countries (Davis, Couper, Janz, Caldwell, & Resnicow, 2010; Harzing, 2006). Developing countries differ from more developed countries in many ways that can affect survey research. For example, the administrative data required for sampling is generally poor and often non-existent (Bulmer, 2001). In addition, non-Westerners are more sensitive to situational variables and this increases the chances and the impact of effects of situational variables in surveys (Peil, 2001; Schwartz, Oyserman, & Peytcheva, 2010). Such variables also affect the respondents' use of RSs (Baumgartner & Steenkamp, 2001; Gibbons, Zellner, & Rudek, 1999). While the problems with administrative data may in part account for the neglect of developing countries in data quality research, generalising research findings about RSs from Western to non-Western contexts is problematic. In spite of the limitations, researchers need to find ways of conducting more data quality research in developing countries since it is by conducting more research that the quality of the required data will also improve.

Culture also affects RSs. This realisation has increased the awareness of the need to control RSs in cross-cultural research (Gibbons et al., 1999; Harzing, 2006; Van Herk, Poortinga, & Verhallen, 2004). However, researchers have generally neglected the possible effect of within-country subcultures on RSs. If culture affects RSs, then it is reasonable to believe that this is not limited to between-country cultures. Within country subcultures should be expected to have similar effects on RSs subject to the size of the cultural distance.

Shorter cultural distances between subcultural groups should result in less substantial RSs effects on data quality, but one cannot argue that such biases are negligible. The absence of bias must be demonstrated and not assumed (Van de Vijver & Leung, 1997). The paucity of research on RSs across within-country subcultures is therefore a limitation of the current RSs literature.

An aspect of subcultural effects on RSs that is worthy of investigation is the rural–urban divide. The debate over the rural–urban divide spans many decades beginning effectively with Wirth's Urbanism Theory (Wirth, 1938) which suggest that urban areas foster more individualism and tolerance of ambiguity than their rural counterparts. If rural and urban areas differ with respect to RSs, then the practice of pooling within-country, survey data across such areas without controlling the RSs may be just as inappropriate and detrimental to data quality as pooling across between-country cultures. Consistent with the stance of demonstrating rather than assuming the absence bias, the existence of within-country rural–urban RSs differentials needs to be investigated. If such RSs differentials exist, the next logical steps are to determine the effect on measurements and on substantive research results.

A major issue in the literature is the identification and measurement of RSs. There are several methods of measuring and controlling RSs and some are implemented with confirmatory factor analysis and latent class analysis (Billiet & McClendon, 2000; Moors, 2003; Weijters, Schillewaert, & Geuens, 2008). Some procedures involve a risk of confounding content with style by measuring both content and style with the same items. Representative indicators approaches to measuring RSs avoid this risk by using a separate set of heterogeneous items from several different content areas to measure the RSs (Greenleaf, 1992).

However, modelling and controlling RSs with representative indicators seem so far to be restricted to confirmatory factor analysis with continuous indicators. In particular, there are no examples in the literature in which RSs are controlled with representative indicators when the data are subjected to latent class analysis or when the manifest variables in factor models are regarded as ordinal. Furthermore, whenever confirmatory factor analysis is employed with representative indicators, the RSs modelled are determined beforehand and researchers tend to default to some combination of the traditionally more recognised RSs (ARS, ERS, DARS and MRS). Given that culture affects RSs, these popular RSs may not be important for all contexts. As a consequence, researchers may not be controlling adequately for the salient RSs.

Both ways of determining the important RSs in a particular context and extensions of representative indicators corrections for RSs to latent class analysis and factor models with categorical indicators are important areas for methodological development. In the process of such development, it is important to establish convergent validity among the methods so that we are assured that they measure the same thing.

This dissertation deals with RSs and the quality of survey data and it addresses each of the issues highlighted in this introduction. It moves from surveying the RSs literature to investigating the effects or RSs in within-country research, to examining the results for RSs between LCA and CFA and finally to demonstrating new methods for correcting for RSs in LCA and CFA with ordinal indicators based on representative indicators approaches. Survey data collected in Guyana are used in the studies conducted and in each case, the RSs are measured with representative indicators. The data are analysed primarily with confirmatory factor analysis and latent class analysis.

The content of this dissertation is organised into self-contained chapters in the sense that each could be read and understood independently. Apart form Chapter 2 and Chapter 9 which present a description of the data used, and the conclusions respectively, each chapter is written as paper that is either published, submitted to a journal or will be submitted in the near future. In total six paper are included in this dissertation.

Chapter 3 — Response Styles in Survey Research: A Literature Review of Antecedents, Consequences and Remedies — which is published in the International Journal of Public Opinion Research, presents the literature review about RSs. This paper discusses the types of RSs, their potential sources and ways to diagnose and control for them. It also identifies several avenues for further research on the topic.

Chapter 4 — Response Styles and the Rural-Urban Divide — presents the second article which is published in Educational and Psychological Measurement. This paper investigates the effect of the rural–urban divide on mean RSs and their relationships with the sociodemographic characteristics of the respondents. It uses the Representative Indicator Response Style Means and Covariance Structure (RIRSMACS) method which implies the use of a confirmatory factor analysis framework. The results of this paper provide answers to the questions about the extent to which within-country subculture affects RSs and about whether culture moderates the effect of the respondents' sociodemographic variables on RSs.

Chapter 5 — Measurement Invariance, Response Styles and Rural-Urban Measurement Comparability — is the third paper. It follows on from the previous chapter to investigate whether the rural–urban RSs differentials affect measurement comparability in Guyana. The paper uses the RIRSMACS model to investigate whether traditional measurement invariance tests provide adequate assurance of the absence of bias across the rural-urban divide. It also provides insights into the effects of RSs on measurement invariance evaluations. In order to do this, configural, metric and scalar invariance are evaluated between rural and urban areas in Guyana with respect to several substantive constructs. This paper is accepted for publication at the Journal of Cross-Cultural Psychology.

Chapter 6 — Measuring Institutional Trust in Guyana: A Second-Order Factor Model with Corrections for Response Styles — is the forth paper and it focuses on evaluating a measurement model for institutional trust with corrections for RSs and on comparing the results of substantive research based on this model with the approaches of using individual items, sum scores and factor models without RSs controlled. In this paper, the RSs are modelled using the RIRSMACS model.

On the one hand, the issue of determining the dimensions of institutional trust in Guyana is addressed and this contributes to the literature on institutional trust in less consolidated democracies. On the other hand, the impact of RSs on structural relationships in within-country research highlighted. Although the impact of RSs is not approached with respect to the rural-urban divide, this paper is a continuation of the theme of demonstrating the importance of RSs in within-country research. In this case, the focus is on effects on the relationships between variables. It also demonstrates the effect of the various methods of measurement — individual items, sum scores, factor analysis and factor analysis with corrections for RSs — on substantive research results and provides guidelines on how to adjust for RSs. This paper will be submitted to a journal in the near future.

Chapter 7 — Are Response Styles Comparable between Latent Class Analysis and Confirmatory Factor Analysis? — compares latent class analysis (LCA) and confirmatory factor analysis (CFA) implementations of representative indicators approaches to modelling response styles (RSs). This paper addresses two main issues. First, CFA researchers tend to default to measuring and controlling some combination of the traditionally more recognised RSs — ARS, ERS, DARS and MRS — because the decision on which styles to include has to be done beforehand. However, the exploratory nature of LCA presents the opportunity to evaluate whether these RSs are salient in the particular context. Second, although RSs may be studied with representative indicators within the LCA framework, there is still a need for extensions of the methodology to cater for corrections for RSs with LCA. This paper investigates the extent of convergent validity between the styles measured by the two techniques and thereby assist in determining whether representative indicators approaches are good candidates for extensions for making adjustments when data are analysed with LCA. This paper will be submitted to a journal in the near future.

Chapter 8 — Factor Mixture Representative Indicators Corrections for Response Styles in Latent Class and Factor Models — is the final paper. It extends representative indicators adjustments for RSs to latent class models and factor models with ordinal indicators. This is done with the use of factor mixture models.

On one hand, it demonstrates the use of a factor mixture model to adjust for RSs in a common factor model that is estimated with categorical indicators by modelling the RSs as latent classifications. On the other hand, it demonstrates the use of a factor mixture model to make RSs adjustments to the measurement of a categorical latent variable which is estimated with latent class analysis. Both approaches to modelling RSs are novel from the perspective that representative indicators adjustments for RSs have previously been restricted to common factor models. The models are presented along with guidelines on how to implement them and how to use them in substantive research. This paper will be submitted to a journal in the near future.

Finally, Chapter 9 — Conclusion — provides a brief conclusion to this dissertation. It summarises the major findings, outlines the limitations and identifies areas for future advancement in relation to RSs research.

References

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*(2), 229–247.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Bulmer, M. (2001). General introduction. In M. Bulmer & D. P. Warwick (Eds.), *Socal research in developing countries: Surveys and censuses in the third world* (pp. 3–24). London: Taylor & Francis Group.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*, 64–73.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26.

Gibbons, J. L., Zellner, J. a., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by spanish-english bilinguals. *Cross-Cultural Research*, *33*(4), 369–381.

Greenleaf, E. A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Krosnick, J. a. (1999). Survey research. *Annual review of psychology*, *50*, 537–567.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, *22*(1), 93–119.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Peil, M. (2001). Situational variables. In M. Bulmer & D. P. Warwick (Eds.), *Socal research in developing countries: Surveys and censuses in the third world* (pp. 71–88). London: Taylor & Francis Group.

Schwartz, N., Oyserman, D., & Peytcheva, E. (2010). Cognition, communication and culture: Implications for the survey response process. In J. A. Harkness et al. (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (pp. 177–190). New Jersey: Wiley & Sons.

Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 143–179). Boston, MA: McGraw-Hill.

Smith, T. W. (2011). Refining the total error perspective. *International Journal of Public Opinion Research*, *23*, 464–484.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications Inc.

Weijters, B. (2006). *Response styles in consumer research* (Unpublished doctoral dissertation). Ghent University.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, *44*(1), 1–24.

CHAPTER 2

DESCRIPTION OF THE DATA

Chapter 2

Description of the Data

2.1 Overview of Survey

2.1.1 Methodological Issues

The data used in this dissertation were are obtained from the Values and Poverty Study in Guyana (VAPO Guyana). This study was set up in the context of an Own Initiative project funded by the Flemish Inter-University Council (VLIR); grant number ZEIN2008PR357, and it was jointly executed by the University of Guyana and Ghent University. It was designed to investigate both methodological and substantive issues and it provides and opportunity to study response styles (RSs) in a non-Western setting. The substantive issues cover a variety of topics including social, cultural, economic and political values and attitudes in regard to society, politics, social inequality and poverty. The interviewers who participated in the study also completed the survey questionnaire. These two groups are linked in the VAPO data and this makes it possible to study interviewers and respondents separately or in combination.

Figure 2.1: Administrative Regions of Guyana[1]

Guyana has an area of 214970 square kilometres and a population of approximately 751223 (Bureau of Statistics, 2002) and it is the only English-speaking territory on the mainland of South America. The country is divided into ten administrative regions (see Figure 2.1). Regions 2, 3, 4, 5, 6 and 10

[1]Obtained from Wikipedia: http://en.wikipedia.org/wiki/Regions_of_Guyana

are regarded as coastal regions and they account for approximately 90% of the total population of the country (Bureau of Statistics, 2002). The VAPO Guyana targeted the adult (age $\geq$ 18 years) population and the survey was executed in two phases. The first phase focused on the coastal regions and it was executed between April and May 2012. The second phase focused on the Hinterland regions (region 1, 7, 8 and 9) and this phase was executed between October and November 2013. Only data from the first phase were available for this dissertation and as such, the remainder of this description is relevant to the first phase of the VAPO Guyana.[2]

The data were collected via face-to-face interviews by a survey organisation (DPMC) under the supervision of the Universities of Guyana and Ghent. The interviewers were trained by DPMC and they attended a two-day briefing organised by the VAPO research team (Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012). The briefing sessions included a general introduction to interviewing which was a refresher for the interviewers; explanations and demonstrations of the contact and selection procedures; and introductions to the content of the questionnaire. A field manual was also provided to guide the interviewers in the event that they encountered difficulties or had questions later.

The sampling procedure included stratification by region (proportional to size), stratification by area type (political demarcations: rural and urban), systematic sampling of the municipalities and cluster sampling of respondents (one per household) within the municipalities. For the systematic and cluster sampling aspects of the selection, the municipalities were arranged within each region in descending order of size. The villages were also arranged in descending order of size within each municipality. At each step of the systematic procedure, 12 households were identified and one adult per household was selected based on a birthday rule. The individual with the next upcoming

[2]For additional information on the VAPO methodology and data summaries, see the VAPO Guyana Codebook.

birthday was the selected respondent. This birthday rule was expected to control for variables such as age, gender and education. The sampling procedure resulted in the selection of 87 clusters within 51 municipalities and in total, 1048 individuals were interviewed (Table 2.1) at a response rate of approximately 87% (American Association for Public Opinion Research, 2011, RR2). To adjust for non-response, the data are weighted by iterative proportional fitting based on the cross-tabulation of age and gender, and separate tabulations of education and voting behaviour.

Table 2.1: Sample Size by Region

Region	Municipalities	Clusters	Realised Sample (n)
2	5	6	72
3	10	13	164
4	15	40	478
5	6	7	80
6	13	16	193
10	2	5	61
Total	51	87	1048

One of the problems involved in conducting surveys in developing countries is the absence or questionable quality of the administrative data required for sampling (Bulmer, 2001). The administrative data for the sampling in the VAPO Guyana were obtained from the 2002 census and these data were approximately ten years old. While the data were thought to be able to indicate the population proportions, listings of the names and addresses of individuals were not used. As an alternative, the VAPO Guyana employed a random walk procedure to identify the households from which the respondents were selected. The procedure was modified depending on the layout and physical features of the areas selected and its execution was closely monitored to ensure that it was implemented as planned and to provide solutions to any problems that were encountered. Using the random walk procedure, attempts were made to contact a total of 1212 potential respondents. Approximately 13% of these contact attempts resulted in nonresponse: 8.3% (refusal), 3.8% (non-contact/

unavailable) and 0.9% (unable to participate due to illness).

2.1.2 The VAPO Guyana Questionnaire

Preparation of the survey questionnaire spanned several months: May 2011 to March 2012. For the RSs component of the questionnaire, forty-five (45) attitude items were selected from various constructs covering several topics (including government, politics, society, crime gender roles and many more). These items were tested in a PAPI survey among students (n=1000) at the University of Guyana leading to the selection of 35 items with low inter-correlations ($|r| \leq 0.3$). The selected items were then included in the VAPO Guyana questionnaire to ensure that separate items are always available to measure RSs in addition to the substantive constructs included (Vander Weyden et al., 2012). Following the identification of the RSs items, the preparation of the items for the substantive topics began along with the construction of the contact form. Many of the questions and item scales were selected from well-developed surveys in Western Europe and Latin America, namely the Belgian National Election Study (BNES); the European Values Study (EVS) and the Americas Barometer (LAPOP) and five of the RSs items were absorbed by some of these scales. However, these items could be easily replaced by making a random selection of one item from the scale. The refinement process focused mainly on phrasing the items for appropriate interpretation in the context and on deleting or replacing items that lack relevance in the Guyanese context. The items included in the VAPO questionnaire are included in Appendix A.3

2.2 Data Summaries

Although a set of items are designated for RSs, they are not necessarily the only items that may be used in this capacity. A random selection of items from various scales in the questionnaire may still be used to supplement the list to RSs items. The full list of items (42) used at one time or another to measure RSs is presented in Table 2.2.

Table 2.2: RSs items

Item	Mean	Standard Deviation
Striving for personal success is more important than providing for good relations with your fellow man	3.75	0.94
I approve of people participating in legal demonstrations	3.37	1.07
In my daily life, I seldom have time to do the things I really enjoy	3.24	1.04
Doctors keep the whole truth from their patients	3.89	0.75
Citizens should spend at least some of their free time helping others	3.98	0.88
Nowadays businesses are only interested in making profits and not in improving service or quality for customers	4.32	0.66
Men should take as much responsibility as women for the home and children	2.90	1.05
I am satisfied with the way democracy works in Guyana	3.48	1.16
When there are children in the home, parents should stay together even if they don't get along	3.24	1.00
I never seem to have enough time to get everything done in my job	2.87	1.13
I am a quiet and shy person	2.96	1.12

All items are scored on a 5-point rating scale: 1 – Completely Disagree, 2 – Disagree, 3 – Neither Agree nor Disagree, 4 – Agree, 5 – Completely Agree

Item	Mean	Standard Deviation
Torturing a prisoner in a Guyanese prison is never justified, even if it might provide information that could prevent a terrorist attack	3.10	1.22
When jobs are scarce, men should have more right to a job than women	4.22	0.63
Schools must teach children to obey authority	3.41	1.03
Employees often pretend they are sick in order to stay at home	3.14	1.13
On the whole, my life is close to how I would like it to be	2.61	1.12
If I help someone, I expect some help in return	4.27	0.74
There are people in my life who really care about me	2.83	1.18
If you want to make money, you can't always act honestly	3.42	1.09
The prison breaks reflect the failure of the judicial system	3.02	1.30
For crimes such as murder and drug traffic, young people from 14 years onwards should be sentenced just as adults	2.72	1.06
Economic growth always harms the environment	3.86	0.73
Participation of citizens in issues concerning the society should be enhanced	3.61	1.01
Guyana is suffering from an economic crisis	3.06	0.96
I trust the media in Guyana	4.01	0.83

All items are scored on a 5-point rating scale: 1 – Completely Disagree, 2 – Disagree, 3 – Neither Agree nor Disagree, 4 – Agree, 5 – Completely Agree

Item	Mean	Standard Deviation
Generally, I am in good health	3.49	0.96
Modern science can be relied on to solve our environmental problems	2.44	1.27
The standard of living of pensioners in Guyana is acceptable	2.52	1.06
The tax authorities are efficient at things like handling queries on time, avoiding mistakes and preventing fraud	3.99	0.88
The Guyanese government, more than the private sector, should be primarily responsible for creating jobs	3.92	0.92
The level of crime that we have now represents a threat to our future wellbeing	3.37	1.05
People like me are being systematically neglected, whereas other groups received more than they deserve	3.25	1.07
I feel myself powerless and at the mercy of current changes	3.18	1.06
These days, you really don't know who you can trust	4.20	0.73
Nowadays, politics has a total lack of common sense	3.38	1.02
Same-sex couples should have the right to marry	1.61	0.92
All politicians are profiteers	3.40	1.04
The parliament does not succeed in solving problems, it is therefore better to abolish it	2.32	1.07
The people should govern directly rather than through elected representatives	2.27	1.02

All items are scored on a 5-point rating scale: 1 – Completely Disagree, 2 – Disagree, 3 – Neither Agree nor Disagree, 4 – Agree, 5 – Completely Agree

Item	Mean	Standard Deviation
The differences between classes ought to be smaller than they are at the present	3.71	0.85
Poverty is a situation in which people are confronted with the negative results of underdevelopment of the country	3.76	0.94
Poverty can only be solved by more equality in international relationships between countries	3.66	0.88
All items are scored on a 5-point rating scale: 1 – Completely Disagree, 2 – Disagree, 3 – Neither Agree nor Disagree, 4 – Agree, 5 – Completely Agree		

Table 2.3: Level of Education

Level of Education	Total	Percentage
Primary or lower	318	30.40
Secondary	598	57.10
More than secondary	131	12.50
Missing	1	0.10

Table 2.4: Ethnicity

Ethnicity	Total	Percentage
Afro-Guyanese/ Black	320	30.60
Amerindian	27	2.60
Chinese	1	0.10
Indo-Guyanese/ East Indian	482	46.00
European/ White	4	0.40
Mixed	213	20.30
Portuguese	2	0.20

The average age of the respondents in phase one of the VAPO Guyana is 36.25 years and Males and females account for 49.20% and 50.80% respectively the sample. A majority of the respondents (57.10%) have attained at most secondary education whereas only approximately 12.50% have attained higher education (see Table 2.3). The largest ethnic group is East Indian. This group accounts for approximately 46% of the population of the coastal regions (See Table 2.4). Rounding out the largest three ethnic groups are the Afro-Guyanese (30.60%) and the group of mixed respondents (20.30%). Each of the other ethnic groups are small. Together, the account for approximately 3.20% of the sample.

References

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Deerfield, IL: AAPOR.

Bulmer, M. (2001). General introduction. In M. Bulmer & D. P. Warwick (Eds.), *Socal research in developing countries: Surveys and censuses in the third world* (pp. 3–24). London: Taylor & Francis Group.

Bureau of Statistics. (2002). 2002 population & housing census. Retrieved from `http://www.statisticsguyana.gov.gy/census.html`

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana*. Ghent, Belgium: Ghent University.

CHAPTER 3

RESPONSE STYLES IN SURVEY RESEARCH: A LITERATURE REVIEW OF ANTECEDENTS, CONSEQUENCES, AND REMEDIES

Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies

Yves Van Vaerenbergh[1,2] **and Troy Devon Thomas**[3,4]

[1]*Human Relations Research Group, Hogeschool-Universiteit Brussel.* [2]*Center for Service Intelligence, Ghent University.*

[3]*Department of Political and Social Sciences, Ghent University.* [4]*Department of Mathematics, Physics and Statistics, University of Guyana*

Abstract

Although the purpose of questionnaire items is to obtain a person's opinion on a certain matter, a respondent's registered opinion may not reflect his or her "true" opinion because of random and systematic errors. Response styles (RSs) are a respondent's tendency to respond to survey questions in certain ways regardless of the content, and they contribute to systematic error. They affect univariate and multivariate distributions of data collected by rating scales and are alternative explanations for many research results. Despite this, RSs are often not controlled in research. This article provides a comprehensive summary of the types of RSs, lists their potential sources, and discusses ways to diagnose and control for them. Finally, areas for further research on RSs are proposed.

Keywords: Response styles, survey research

Published as:

Van Vaerenbergh, Y. & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences and remedies. *International Journal of Public Opinion Research,* 25, 195–217.

3.1 Introduction

In several social sciences disciplines, questionnaire data are indispensable sources of information. Researchers rely on respondents' self-reports to understand their attitudes and behaviours. A popular way to measure these attitudes and behaviours is to use rating scales (Moors, 2010). However, after respondents have provided their ratings for given statements, the question of whether the given answers reflect their true opinions remains.

Researchers agree that a response variance can be decomposed into true and error variances (Smith, 2011), the latter of which includes variance due to response styles (RSs). Thus, RSs distort research results. RSs are the respondent's systematic tendency to respond to a range of survey items on a different basis from what the items are designed to measure (Paulhus, 1991). RSs are present in the entire data set and they affect the validity of research conclusions in two main ways (Baumgartner & Steenkamp, 2001). First, RSs affect univariate distributions, that is, RSs have an impact on means and variances. For example, previous research has typically found gender differences in passive/laissez-faire leadership. However, Moors (2012) finds that women are more likely to use the highest and the lowest response categories of a rating scale (extreme RS) than men, which introduces systematic error into the research results. Consequently, the relationship between gender and leadership styles is spurious when taking RSs into account. Thus, without controlling for RSs, researchers might draw incorrect conclusions from comparative tests such as t-tests or F-tests (Cheung & Rensvold, 2000). Second, RSs affect multivariate distributions. For example, Baumgartner and Steenkamp (2001) correlate health consciousness (HCO), quality consciousness (QCO), environmental consciousness (ECO), and ethnocentrism (ETN) and find the following correlations: HCOQCO: 0.40; HCOECO: 0.33; QCOECO: 0.31; HCOETN: 0.28; QCOETN: 0.19; and ECOETN: 0.15. From a theoretical perspective, one might assume that significant correlations exist among HCO, QCO, and ECO, but not with ETN. However, controlling for RSs substantially reduces the correlations to the following: HCOQCO: 0.20; HCOECO: 0.15; QCOECO: 0.13; HCOETN: 0.02; QCOETN: 0.00; and ECOETN: 0.01. Thus, RSs affect the magnitude of correlations between variables. Many statistical techniques, such as Cronbach's α, regression analysis, factor analysis, and structural equation modelling, rely on correlations between variables (Reynolds & Smith, 2010). As a result, studies examining such relationships without controlling for RSs might yield misleading results.

Therefore, RSs potentially affect all empirical studies that use rating scales and are alternative explanations for the results. However, despite its importance, many researchers do not control for this source of bias. The purpose of this article is to provide insights into RSs by (1) defining different types of RSs, (2) discussing the different sources of RSs, and (3) providing an overview of various statistical remedies for RSs. This is important because, to our knowledge, no comprehensive discussion of RSs is available in the literature. Given that only a few research articles control for RSs, this article provides the necessary background and tools for researchers to assess RSs in their own research projects.

3.2 Types of RSs

The literature distinguishes between several types of RSs. Table provides an overview of eight RSs that are prominent in the literature. Included are acquiescence RS (ARS), disacquiescence RS (DARS), mid-point response style (MRS), extreme response style (ERS), mild response style (MLRS), net acquiescence response style (NARS), response range (RR), and noncontingent response style (NCRS), along with short descriptions, graphical representations when applicable, an overview of the main consequences, and sources from which further explanations can be obtained.

Table 3.1: Types of RSs

Type	Definition	Respondent's use of a 7-point rating scale[a]	Consequences	Representative studies
ARS	Tendency to agree with items highest response categories are used	○○○○●●●	Inflates observed means, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Greenleaf (1992b)
DARS	Tendency to disagree with items regardless of content, only the lowest response categories are used	●●●○○○○	Deflates observed means, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Stening and Everett (1984)

Note. [a]A 7-point scale is used only for illustrative purposes; RSs are also present in other types of rating scales. Black dots indicate the response categories a respondent is more likely to use under a certain RSs.

ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style; MLRS=Mild response style; NARS=Net acquiescence response style; RR=Response range; NCR=Noncontingent responding.

Type	Definition	Respondent's use of a 7-point rating scale[a]	Consequences	Representative studies
MRS	Tendency to use the middle response category of a rating scale, regardless of content	○○○●○○○	Brings observed means closer to the mid-point, deflates variance, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Weijters, Schillewaert, and Geuens (2008)
ERS	Tendency to use the highest and lowest response categories of a rating scale	●○○○○○●	Inflates (deflates) observed means variance, decreases magnitude of multivariate relationships Baumgartner and Steenkamp (2001); Greenleaf (1992b)	

Note. [a]A 7-point scale is used only for illustrative purposes; RSs are also present in other types of rating scales. Black dots indicate the response categories a respondent is more likely to use under a certain RSs.

ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style; MLRS=Mild response style; NARS=Net acquiescence response style; RR=Response range; NCR=Noncontingent responding.

Type	Definition	Respondent's use of a 7-point rating scale[a]	Consequences	Representative studies
MLRS	Tendency to avoid the highest and lowest response categories of a rating scale. This is the complement of ERS	○●●●●●○	Brings observed means closer to the mid-point, deflates variance, increases magnitude of multivariate relationships	Hurley (1998); Moors (2008)
NARS	Tendency to show greater acquiescence than disacquiescence.		Inflates variance, deflates observed means if negative	Baumgartner and Steenkamp (2001); Weijters, Cabooter, and Schillewaert (2010)

Note. [a]A 7-point scale is used only for illustrative purposes; RSs are also present in other types of rating scales. Black dots indicate the response categories a respondent is more likely to use under a certain RSs.

ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style; MLRS=Mild response style; NARS=Net acquiescence response style; RR=Response range; NCR=Noncontingent responding.

Type	Definition	Respondent's use of a 7-point rating scale[a]	Consequences	Representative studies
RR	Tendency to use a narrow or wide range of response categories around the mean response		When large: inflates variance, decreases magnitude of multivariate relationships	Greenleaf (1992b)
NCR	Tendency to respond to items carelessly, randomly, or nonpurposefully		No a priori hypotheses about the effect can be specified	Baumgartner and Steenkamp (2001); Watkins and Cheung (1995)

Note. [a]A 7-point scale is used only for illustrative purposes; RSs are also present in other types of rating scales. Black dots indicate the response categories a respondent is more likely to use under a certain RSs.

ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style; MLRS=Mild response style; NARS=Net acquiescence response style; RR=Response range; NCR=Noncontingent responding.

As Table I indicates, RS have various influences on observed means and/or variances and on the magnitude of the relationships between variables. Researchers have devoted attention mainly to investigating ARS, DARS, ERS, and MRS (Cabooter, 2010; Weijters, 2006). In the remainder of this article, we not only focus on these four types, but also elaborate on other types when necessary.

3.3 Sources of RSs

Weijters (2006) classifies sources of RSs into two main categories: the stimulus level and the respondent level. At the stimulus level, RS are viewed as a consequence of the survey instrument. At the respondent level, RSs are viewed as a consequence of personal characteristics. (Baumgartner & Steenkamp, 2001) note that situational factors can encourage or discourage people's inherent tendency to use RSs. Although we discuss stimulus (situational) and respondent factors separately in the subsequent section, it should be kept in mind that these factors cannot be viewed as independent of each other.

3.3.1 Stimuli as Sources of RSs

According to (Maxey & Sanford, 1992, p. 295) "It seems almost impossible to escape the possibility that questionnaire items influence the responses given by respondents." This suggests that questionnaire design and questionnaire items themselves act as stimuli to respondents, and therefore they may also influence RSs. Table summarizes research on stimuli as sources of RSs. These stimuli include scale format, mode of data collection, cognitive load, interviewer effects, survey language, and topic involvement.

Table 3.2: Stimuli as sources of RSs

Source	ARS	DARS	ERS	MRS	Representative studies
Scale format			One-stage=two-stage scale formats		Arce-Ferrer (2006)
			Two-stage>one-stage scale formats		Albaum, Roster, H, and Rogers (2007)
	Weak evidence of ARS in 5-, 6-, 7-, 9-, 10-, and 11-point scales		No difference between 5-, 6-, 7-, 9-, 10-, and 11-point scales		Kieruj and Moors (2010, 2013); Moors (2008)
	Longer scales have no effect on NARS		Longer scales lead to lower levels of ERS		Weijters, Cabooter, and Schillewaert (2010)
	Neutral point leads to higher levels of NARS		Neutral point leads to lower levels of ERS		

Note. ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style.

Source	ARS	DARS	ERS	MRS	Representative studies
	Fully labelled scales increase ARS		Fully labelled scales reduce ERS		
Mode of data collection	Telephone>face-to-face		Telephone>face-to-face		Jordan, Marcus, and Reeder (1980)
	Telephone>paper and pencil and web	Paper and pencil>Web	Paper and pencil>web	Telephone<Paper and pencil and Web	Weijters et al. (2008)
			Web>paper and pencil		Kiesler and Sproull (1986)
	Web=face-to-face				Heerwegh (2009)
Cognitive load	ARS increases with cognitive load				Knowles and Condon (1999)

Note. ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style.

Source	ARS	DARS	ERS	MRS	Representative studies
	NARS increases with cognitive load				Cabooter (2010)
Interviewer experience	Higher with experienced interviewers				Olson and Bilgen (2011)
	No interviewer effects				Hox, de Leeuw, and Kreft (1991)
Survey language	Native language>second language		Native language>second language	Second language>native language	Harzing (2006)
				Second language>native language	Gibbons, Zellner, and Rudek (1999)

Note. ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style.

Source	ARS	DARS	ERS	MRS	Representative studies
Topic involvement			Increases with higher levels of topic involvement		Gibbons et al. (1999)

Note. ARS=Acquiescence response style; DARS=Disacquiescence response style; MRS=Mid-point response style; ERS=Extreme response style.

Scale format. Greenleaf (1992a) and Baumgartner and Steenkamp (2001) suggest examining RSs for different scales formats, and some researchers have responded. For example, Kieruj and Moors (2010) find that MRS emerges when nine or more response categories are offered, and Kieruj and Moors (2013) find weak evidence of ARS in 5- to 11-point rating scales. Added to this, Weijters, Cabooter, and Schillewaert (2010) find that longer rating scales have no effect on NARS but that NARS increases with the addition of a neutral point and with fully labelled scales. For ERS, the evidence is mixed. Arce-Ferrer (2006) finds no difference in ERS between one- and two-stage rating scales, whereas (Albaum et al., 2007) find higher ERS in two-stage than in one-stage rating scales. One-stage scales are simple scales, whereas two-stage scales have a more in-depth question following an initial filter question. Researchers examining the impact of scale format on ERS have focused mainly on one-stage rating scales. Kieruj and Moors (2010, 2013) compare 5- to 11-point rating scales and find no effect of the number of response categories on ERS, but Weijters, Cabooter, and Schillewaert (2010) use 4- to 7-point scales and find that ERS decreases as the number of response categories increases. Whereas Kieruj and Moors use latent-class confirmatory factor analysis (LCFA) to model ERS, Weijters, Cabooter, and Schillewaert use representative indicators of RS (RIRS). In addition, Kieruj and Moors label only the endpoints of the scales, whereas Weijters, Cabooter, and Schillewaert contrast fully labelled and endpoint-labelled scale formats and find that fully labelled scales reduce ERS. This potentially explains why Kieruj and Moors find no differences in ERS across scale formats.

According to Weijters, Cabooter, and Schillewaert (2010), the optimal number of response categories depends on the purpose for which the scale is to be used. If a researcher wants to report direct summaries of responses, such as means or percentages, Weijters, Cabooter, and Schillewaert (2010) suggest the use of fully labelled 5-point (or 7-point) scales because labelling makes the scale more directly interpretable. This recommendation coincides with that of Krosnick (1999), who contends that fully labelled formats maximize reliability and validity because the labels clarify the meaning of the scale. If instead the researcher wants to relate variables or estimate linear models, Weijters, Cabooter, and Schillewaert (2010) suggest that the endpoint-labelled 5-point (or 7-point) rating scale is best because respondents use such scales in a way that conforms better to linear models. Response scales have also been examined with many other criteria — for example, reliability, information recovery, distribution of scale means, and ease of use (Preston & Colman, 2000; Weng, 2004)— resulting in similar recommendations with respect to the optimal number of response categories.

Modes of data collection. Differences in RSs among modes of data collection lead to important implications for researchers. Telephone surveys lead to higher ARS and ERS and lower MRS than face-to-face, paper-and-pencil, and web surveys (Jordan et al., 1980; Weijters et al., 2008). These findings suggest that the mode of data collection influences research results even when only one mode of data collection is used. If researchers use telephone surveys, they should interpret raw mean scores and variances cautiously. Mode effects on RS are also important in light of the increased popularity of mixed-mode surveys (Heerwegh, 2009). Researchers using mixed-mode data collections should be cautious about combining data coming from different modes because RSs might induce observed differences in the results. Therefore, researchers should account for RSs in the analysis of mixed-mode data.

Cognitive load. To our knowledge, only two studies have focused on the relationship between cognitive load and RSs. Knowles and Condon (1999) find that ARS increases with cognitive load, and Cabooter (2010) finds that NARS increases with cognitive load. Cognitive load is present in many situations, and researchers should try to avoid it. Researchers can do so by inviting respondents to participate in lab research, allowing respondents to participate when they have time available, or providing a room where they can relax, to name a few (Cabooter, 2010). Researchers should also word survey questions clearly, as suboptimal question wording requires more cognitive effort to understand the meaning of the questions (Lenzner, Kaczmirek, & Galesic, 2011). If researchers suspect that the respondents completed a survey under high cognitive load, they should conduct a post hoc assessment of RSs.

Interviewer effects. Interviewer effects on RSs have received limited attention in the literature. Olson and Bilgen (2011) find that experienced interviewers influence higher levels of ARS, but Hox et al. (1991) find no such effect. Despite the potential effect of interviewer experience on ARS, in general, experienced interviewers decrease measurement errors from other sources, such as nonresponse (Lipps & Pollien, 2011)or social desirability (Cleary, Mechanic, & Weiss, 1981). Experienced interviewers are therefore preferred, but researchers should still control for RSs.

Survey language. In general, researchers should adapt questionnaires to the local language (Usunier, 2011); however, administering questionnaires in a second language leads to lower levels of ARS and ERS but higher levels of MRS and RR than when administered in a native language (Gibbons et al., 1999; Harzing, 2006). Overall, respondents make better use of the entire scale when responding to surveys in their native language, instead of mainly using the scale's mid-point. These findings are important because cross-cultural studies often administer questionnaires in English across different language groups (Rowland et al., 2010). Preferably, respondents should complete surveys in their native language because they are better able to qualify their answers on rating scales. Nevertheless, a post hoc assessment of ARS and ERS is necessary.

Topic involvement. Although topic involvement is perhaps more a task characteristic than a stimulus, we consider it because it is related to the content of the question. If an item or question is not relevant to a respondent, there will be lower involvement, which influences RSs. For example, Gibbons et al. (1999) report that ERS is more prevalent if the respondent is more involved with the presented stimulus.

3.3.2 Respondents as Sources of Response Styles

Researchers who subscribe to the view that RSs are due to the respondent argue that RSs are mainly determined by the respondent's characteristics and personality. We first consider demographic variables and then explore personality and culture.

Education. With few exceptions, research indicates that education is inversely related to RSs. Meisenberg and Williams (2008) find this to be nearly a worldwide phenomenon for ARS and ERS. However, research findings are not unanimous, and not all RSs are investigated by each researcher. Weijters, Geuens, and Schillewaert (2010b) focus on ARS, DARS, ERS, and MRS and, except for DARS, find inverse relationships to education. However, Moors (2008) and De Jong, Steenkamp, Fox, and Baumgartner (2008) find no effect of education on ERS.

Matarazzo and Herman (1984) indicate that education is correlated with IQ and suggest that in cases of extreme absence of data, the level of education can be used as an indicator of IQ. Therefore, some link exists between education and IQ. For ERS, Light, Zax, and Gardiner (1965) find a negative relationship with IQ. In addition, they find lower MRS among older people with high IQ than younger people with high IQ but find the reverse for low-IQ people. In this case, the ages included ranged from 9 to 18 years. In addition, with intelligence measured by the American College Exam, Zuckerman and Norton (1961) find that ARS decreases as intelligence increases.

Age. Research has also questioned whether a relationship exists between age and RSs (Stukovsky, Palat, & Seldlakova, 1982). For ARS, research shows evidence of a positive relationship with age (Billiet & McClendon, 2000; Greenleaf, 1992a; Ross & Mirowsky, 1984; Weijters, Geuens, & Schillewaert, 2010b), but Eid and Rauber (2000) report no effect. The evidence for ERS is particularly interesting. Several researchers find that ERS increases with age (Greenleaf, 1992b; Meisenberg & Williams, 2008; Weijters, Geuens, & Schillewaert, 2010b), others find that older respondents have lower levels of ERS (Austin, Deary, & Egan, 2006; Light et al., 1965), and still others find no effect (Johnson, Kulesa, Cho, & Shavitt, 2005; Moors, 2008). However, De Jong et al. (2008) find that both the younger and the elderly respondents have higher levels of ERS than the middle-aged group. This curvilinear relationship potentially explains the different findings. For example, if there is a higher proportion of elderly respondents than younger and middle-aged respondents and elderly respondents have higher ERS, one might assume a positive linear relationship between age and ERS. Conversely, if the proportion of younger respondents is higher and the younger respondents have higher ERS, a negative linear relationship with age might be assumed. Alternatively, if the proportions of younger and elderly respondents are about equal and the two groups both have higher ERS, linear modelling should find no effect. For DARS and MRS, Weijters, Geuens, and Schillewaert (2010b) find no effect and a positive relationship for age, respectively.

Gender. Some studies report higher ARS for women than men (Austin et al., 2006; Weijters, Geuens, & Schillewaert, 2010b), whereas others report no gender effect (Light et al., 1965; Marin, Gamba, & Marin, 1992). For ERS, the results include a greater tendency among women (De Jong et al., 2008; Weijters, Geuens, & Schillewaert, 2010b), a greater tendency among men (Harzing, 2006; Meisenberg & Williams, 2008), and no gender effect (Grimm & Church, 1999; Light et al., 1965; Marin et al., 1992; Moors, 2008). For DARS, Crandall (1973) finds no relationship with gender. For MRS, Harzing (2006) finds higher levels among women, but Light et al. (1965) and Grimm and Church (1999) find no gender effect.

Income and employment. In general, ARS and ERS are higher when socio-economic status and income are lower (Greenleaf, 1992a, 1992b; Meisenberg & Williams, 2008; Ross & Mirowsky, 1984). In addition, Johnson et al. (2005) indicate that length of employment is positively related to ARS but not to ERS. Contrary to the latter, Eid and Rauber (2000) find a positive relationship between length of employment and ERS.

Race. Prior research has found that race is a significant antecedent of RSs. For example, some studies indicate that African Americans and Hispanics exhibit higher levels of ARS and ERS than White Americans (Bachman & O'Malley, 1984; Marin et al., 1992). Baron-Epel, Kaplan, Weinstein, and Green (2010) also report that ERS and MRS are higher among Jews than Arabs in Israel. These findings suggest that RSs might be higher among minority groups. However, Naemi, Beal, and Payne (2009) find no support for such a conclusion regarding ERS.

In general, the literature indicates that socio-demographic variables affect RS, which suggests that researchers should be careful when comparing results across demographic profiles. However, the findings are not always consistent. A potential explanation is that empirical findings on the relationships between socio-demographic variables and RS are mere reflections of personality (Moors, 2008).

Personality. Support for the stability and consistency of RSs in the literature, stability throughout the questionnaire (Hamilton, 1968, in relation to ERS), consistency throughout the questionnaire (Naemi et al., 2009; Weijters, Geuens, & Schillewaert, 2010a, in relation to ARS and ERS), stability between data collections with a 1-year time gap (Weijters, Geuens, & Schillewaert, 2010b, in relation to ARS, DARS, MRS and ERS), and stability over a 4-year period with the same respondents (Billiet & Davidov, 2008, in relation to ARS) might be enough to counter Rorer's (1965) rejection of the notion that personality affects RS. In addition, previous research has found that ERS is positively related to intolerance of ambiguity (Brengelmann, 1960; Naemi et al., 2009), preference for simple thinking and decisiveness (Naemi et al., 2009), and the Big Five personality traits extraversion and conscientiousness (Austin et al., 2006). Furthermore, Ayidiya and McClendon (1990) report a positive relationship between MRS and evasiveness, and Couch and Keniston (1960) find that ARS is positively related to impulsiveness and extraversion.

However, all previous findings on the role of personality have been criticized because rating scales are used to assess personality, and thus the personality measures themselves might be contaminated with RSs (Bentler, Jackson, & Messick, 1971). Naemi et al.'s (2009) attempt to let a close friend complete the personality measures does not overcome this limitation. Conversely, Cabooter (2010) investigates "self-regulatory focus" and ERS and MRS with the use of unique, scale-free personality measures and finds that a prevention focus is positively related to MRS and a promotion focus is positively related to ERS. These findings validate the existence of relationships between personality and RSs, but because nearly all research has focused on ERS, our understanding is limited to this RSs.

That personality predicts RSs behavior makes it difficult, if not impossible, for researchers to prevent respondents' use of RSs (Kieruj & Moors, 2010). Therefore, researchers should diagnose and correct for RSs.

Culture- and country-level characteristics. Many studies highlight the relationship between RSs and cultural (or cross-national) differences. Clarke III (2000) finds the main effect of culture on ERS and indicates that ERS varies across countries and across subcultures within countries. Meisenberg and Williams (2008) report that countries with low-IQ levels show higher ERS, that countries with corrupt societies show both higher ERS and ARS, and that democracy and political freedom do not affect ARS and ERS. Van Herk, Poortinga, and Verhallen (2004) find that Mediterranean countries (Greece, Italy, and Spain) have higher ARS and ERS than Western European countries

(England, Germany, and France). They also conclude that ARS and ERS increase as individualism — one of Hofstede's dimensions — decreases. However, they do not include all of Hofstede's dimensions (individualism, uncertainty avoidance, masculinity, and power distance; see Hofstede, 2001), and since the groups of countries may vary on the other dimensions, the effect of individualism is not unequivocally established. Grimm and Church (1999) find no consistent effect of individualism on ARS or ERS and no effect of culture on MRS, whereas Johnson et al. (2005) find that the four dimensions are each negatively related to ARS and that power distance and masculinity are positively correlated with ERS. In addition to this, De Jong et al. (2008) find a positive relationship between ERS and individualism, uncertainty avoidance, and masculinity, whereas Chen, Lee, and Stevenson (1995) report a negative relationship between MRS and individualism.

Harzing (2006) examines the effect of RSs on cultural variables by including both Hofstede's variables and variables based on the GLOBE dimensions (House, Hanges, Javidan, Dorfman, & Gupta, 2004, see). Harzing uses the GLOBE values for power distance, in-group collectivism, institutional collectivism, and uncertainty avoidance in two categories (values, or "what should be," and practices, or "what is"), resulting in eight variables. The findings indicate that both the nature of the relationships (whether positive or negative) and whether the relationships can be generalized (statistical significance) sometimes depend on the method of calculation (Hofstede or GLOBE values).

The relationships between culture and RSs have important implications for cross-cultural (or cross-national) research. Given that obtained means, variances, and covariances are biased by RSs (Baumgartner & Steenkamp, 2001), traditional measurement equivalence tests should be corrected for RSs. For example, Welkenhuysen-Gybels, Billiet, and Cambré (2003) and Kankaraš and Moors (2011) demonstrate that the results of measurement equivalence tests can change substantially when adjustments are made respectively for

ARS and ERS.

Overall, demographic and personality variables explain a relatively small proportion of the variance of RSs, whereas culture and country-level characteristics seem to explain a relatively large proportion of RSs in cross-cultural studies. Using a Belgian sample, Weijters, Geuens, and Schillewaert (2010b) find that demographic variables explain between 1.4% and 8.3% of the variance in RSs depending on which RS is considered, whereas Meisenberg and Williams (2008) find that socio-demographic variables (e.g., corruption, gross domestic product) explain 15% of the variance in ARS and ERS at the individual level but that country characteristics explain 63.2% (ARS) to 74.5% (ERS) at the country level. In addition, De Jong et al. (2008) indicate that Hofstede's dimensions explain 59% of the between-country variance in ERS. However, because Hofstede and McCrae (2004) find significant and substantial correlations between each of Hofstede's cultural dimensions and personality (specifically, extraversion, conscientiousness, openness to experience, neuroticism, and agreeableness, as measured by the revised NEO personality inventory), overlap occurs between personality and culture. It is therefore, not clear whether the indicated explanatory power for culture represents the unique effect of culture. Furthermore, although socio-demographics explain the smaller proportion of the variance in RSs, they are still important determinants of RSs. The effect of the personal antecedents varies from study to study, and so the explanatory power also likely varies. Neglecting socio-demographic variables as a means of controlling for RSs when the data differ in relation to demographics is potentially damaging to research.

3.4 Diagnosing and Remedying RSs

The literature identifies several ways to diagnose and control RSs. Table 3.3 provides an overview of the different approaches. In comparing the different techniques, several remarks are appropriate.

Table 3.3: Methods of detecting and correcting for RSs

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
Count procedure	Count the number of agreements, disagreements, extreme responses, and/or mid-point responses on substantive measures across an entire questionnaire	Easy to use, no additional indicators are necessary	Only works with heterogeneous items	Bachman and O'Malley (1984); Reynolds and Smith (2010)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
Counting double agreements on reversed items	Include reversed items in the questionnaire, and count the number of double agreements on the reversed items	Easy to use, no additional indicators are necessary	Sometimes difficult to formulate reversed items, people's responses to reversed items might be due to interpretational factors	Hox et al. (1991); Johnson et al. (2005)
MTMM	The same trait is repeatedly measured by means of different methods. Observed variance can be decomposed into true variance and error variance	Easy to set up, easy to use, measures net effects of ARS and DARS, no additional indicators are necessary	Gives no indication of ERS and MRS, consistency bias and memory effects might arise due to repeated measurement, problems of identification arise often	Saris, Satorra, and Coenders (2004a); Saris, Satorra, and Coenders (2004b)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
Specify method factor in CFA	Specify positive and negative loadings on content factor, specify positive loadings on a method factor	Relatively easy to specify, most researchers are familiar with CFA, no additional indicators are necessary	Does not control for DARS, MRS, or ERS; requires the use of balanced scale items; all loadings on the method factor are restricted to equality in order to identify the model	Billiet and McClendon (2000); Welkenhuysen-Gybels et al. (2003)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
Latent-class regression analysis	Run a latent-class regression analysis, and assess whether a method factor emerges	No additional indicators are necessary	Specific software is necessary, researchers might be unfamiliar with latent-class analysis, sometimes hard to specify	Moors (2010); Van Rosmalen, van Herk, and Groenen (2010)
LCFA	Specify two method factors, one to measure ARS, one to measure ERS	No additional indicators are necessary, recent models allow discriminating ARS and ERS	Does not account for DARS and MRS, specific software is necessary, researchers might be unfamiliar with LCFA	Moors (2003, 2012); Kieruj and Moors (2010, 2013)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
IRT model	Models the probability of ticking a certain response option as a function of the underlying latent variable	Allows different items to be differentially useful for measuring ERS, relaxes the assumption that ERS measures should be uncorrelated	Only developed for ERS, requires use of Markov Chain Monte Carlo procedures, which might be more difficult to implement	Bolt and Newton (2011); De Jong et al. (2008)
RIRS method	Include a number of uncorrelated, maximally heterogeneous measures in content to the survey, and calculate weighted RSs indicators	Easy to calculate, allows measuring ARS, DARS, ERS, MRS, NARS, not related to content, easy to include as covariates in subsequent analyses	Additional items need to be added to the survey	Baumgartner and Steenkamp (2001); Greenleaf (1992a, 1992b); Weijters (2006)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
RIRMACS method	Add additional, uncorrelated items to the survey, which serve as observed variables in a CFA; ARS, DARS, MRS and ERS serve as latent variables; Extends the RIRS method	Easy to use, RSs indicators can be added as covariates in subsequent analyses, use of specific RSs indicators allows discrimination between content and style, allows measurement of ARS, DARS, MRS, and ERS; allows testing of convergent and discriminant validity of the different RSs	Additional items need to be added to the survey	Weijters et al. (2008)

Note. MTMM=Multi-traitmulti-method models; CFA=Confirmatory Factor Analysis; LCFA=Latent-class confirmatory factor analysis; IRT=Item-response theory; RIRS=Representative indicators for response styles; RIRMACS=Representative indicators response styles means and covariance structure

First, counting double agreements on reversed items (Johnson et al., 2005), or specifying a method factor on balanced-scale items (Billiet & McClendon, 2000), requires the use of balanced-scale items. This may be problematic, because it is often difficult to formulate reversed items (Billiet & McClendon, 2000) and because the way people respond to reversed items may be due to interpretational issues rather than ARS (Wong, Rindfleisch, & Burroughs, 2003). For example, respondents tend to minimize retrieval of additional information when answering nearby nonreversed items but tend to maximize retrieval of new and different information when answering nearby reversed items (Weijters, Geuens, & Schillewaert, 2009). As a result, balanced scales introduce several other problems that may affect the validity of the research results. Moreover, the majority of measurement scales are not balanced (Baumgartner & Steenkamp, 2001), so these techniques may not always be applicable.

Second, not all approaches simultaneously account for multiple types of RSs. Multi-traitmulti-method models account for ARS and DARS but not ERS or MRS (Saris et al., 2004a). The balanced-scale method (Billiet & McClendon, 2000) accounts only for ARS, while the most recently developed LCFA approach (Kieruj & Moors, 2010, 2013) allows for detection and control of ARS and ERS. The most comprehensive way to detect and control RSs to date is to add RIRSs to the questionnaire, which allows for calculation of ARS, DARS, ERS, MRS, and NARS (Baumgartner & Steenkamp, 2001; Weijters et al., 2008). In regular studies, five items per response style indicator should be included, but in studies explicitly focusing on RS, 1014 items per RS indicator is recommended (Weijters et al., 2008). This may not always be possible because of survey length restrictions.

Third, convergent validity between methods is not well established. De Beuckelaer, Weijters, and Rutten (2010) compare the RIRS method with the more traditional method in which survey items used for substantive purposes are also used to model ARS and ERS (count procedure). The proportion of

ARS is the same for the two methods, but the correlation between the methods is low to very low. In contrast, the proportion of ERS is higher with the traditional method than the RIRS method, but the correlation between the methods is moderate to strong. Convergent validity is therefore not established between the two methods. Kieruj and Moors (2013) also examine convergent validity by correlating a latent class factor, designed to measure ERS, with a RIRS measure of ERS. The two measures of ERS are moderately correlated, thus providing preliminary evidence of convergent validity between the methods, but additional research on this issue is necessary.

To control for RSs, we recommend the use of the RIRS or representative indicators response styles means and covariance structure (RIRMACS) method. These methods enable tests for various types of RSs and the use of RSs as covariates in subsequent analyses. Moreover, the RIRMACS method allows for evaluation of convergent and discriminant validity between the various RSs. Researchers may not always have the means to include additional questions in the survey, may be working on secondary data, or may not want to assume that rating scale data are continuous. In these cases, the LCFA approach provides an alternative. It allows for separation of item content from RSs and does not assume interval level data, and at least for ERS, preliminary evidence of convergent validity with the RIRS method has been established. However, given the uncertainty of convergent validity across methods, researchers should use multiple methods to account for RSs and to assess the stability of their findings across the methods.

3.5 Conclusion

Although the RSs may vary across situational and personal variables, a careful examination of the literature suggests that RSs are often a serious threat to the validity of research results. Because they affect univariate and multivariate distributions, RSs are alternative explanations of most research findings. We contend that researchers should do whatever they can to control for RSs, to obtain more accurate results. Doing so requires both careful examination of the context in which the research is conducted, alongside the tools used to collect data, and the use of statistical procedures to detect and control RSs. Furthermore, we provide an overview that researchers can use when evaluating the potential biasing effects of RSs in their own research projects.

Although researchers have gained substantial knowledge on RSs, not all the issues about this important topic have been resolved, and more work is necessary to enhance understanding of this phenomenon. Next, we provide several suggestions for further research.

3.6 Directions for Further Research

Although RSs have received extensive attention, more work is necessary to extend and improve understanding of its antecedents. First, many conflicting results have emerged in the literature. Therefore, a meta-analysis that examines methodological between-study variables to provide a quantitative assessment of the different findings is necessary. For example, researchers have found differences between ad hoc measures and representative indicators as measures of RSs (De Beuckelaer et al., 2010), and this potentially explains the different findings in the literature.

Second, researchers should also examine the mediating variables between antecedents and RSs. Such examination would provide insights into the cognitive processes underlying the relationships between the antecedents and RSs. Currently, such studies are scarce (Olson & Bilgen, 2011), and thus more work remains to be conducted in this area.

Third, the adverse impact of RSs on research results has recently been demonstrated. For example, Moors (2012) shows that the previously accepted relationship between gender and leadership styles is spurious when RSs are taken into account. Similar work is necessary to convince researchers about the potential consequences of not controlling for RSs and to update existing theories within the various fields.

Regarding the antecedents of RSs, research has focused on investigating either stimulus-related or person-related variables (Weijters, 2006). However, Baumgartner and Steenkamp (2001) note that a person-related source of RSs (e.g., personality) may trigger or attenuate the effects of stimulus-related sources. Research should therefore examine interaction effects among antecedents.

Because we do not yet fully understand how research designs can trigger or retard the use of RSs, further research on stimulus-related antecedents would be useful. Kieruj and Moors (2013) propose that survey length might trigger ARS, but research has not yet formally examined this issue. Naemi et al. (2009) find that the amount of time a respondent spends on the questionnaire significantly influences RSs, and Cabooter (2010) investigates cognitive load (as time pressure) as a situational determinant of RSs. However, other situation-related variables, such as mood, fatigue, or ego depletion, may affect RSs, but these relationships have not been tested properly to date.

Research seems to focus on certain scale formats, and thus several opportunities for further research exist. First, it might be useful to examine culture as a moderator of the scale format–RSs relationship. This would lead to identification of the scale format that suffers least from RSs and which would be of substantial benefit to cross-cultural (or cross-national) research. Second, researchers could examine whether adding a “don’t-know” option to the survey affects RSs. Third, Tourangeau, Couper, and Conrad (2007) examine the impact of scale colour in a web survey on mean responses to a rating scale. They find that for endpoint-labelled scales, when the end points are shaded in different hues compared with the same hue, responses shift toward the high end of the scale. Research should formally examine the impact of different scale colours on RSs. Research might also examine how background colours of a web survey (e.g., colour of banners, background colour itself) influence RSs. Fourth, research could also assess differences in RSs between unipolar and bipolar scales and between other scale formats, such as numbered and unnumbered. Tourangeau et al. (2007) indicate that the effect of shading on mean responses disappears with fully labelled scales and reduces with fully numbered scales, so there might be merit in evaluating numbered and unnumbered scales in relation to RSs. Preferably, researchers should examine all these issues in a factorial design to obtain a comprehensive picture of how scale format influences RSs.

In relation to person-related variables, researchers should further explore the role of personality on RS using scale-free personality tests, such as the one Cabooter (2010) developed. In addition, researchers should either use personality measures that do not overlap with culture (as Harzing, 2006, attempted for extraversion) or explicitly model the joint effect of personality and culture on RSs to quantify the overlap, clarify the unique effect of personality, and provide improved estimates of the explanatory power of culture for RSs.

Another important area for research is RSs measurement. Only a few studies have examined the convergent validity of RSs measures, though various methods have been proposed in the literature (see Table 3.3). Research should further examine convergent validity between methods, preferably through simulations. This can lead to determination of the best (or optimal) method of detecting and/or controlling RSs. In addition, research has recently proposed instructional manipulation checks to detect satisficing (Oppenheimer, Meyvis, & Davidenko, 2009). Research could thus examine the relationships between these instructional manipulation checks and RSs.

Traditional measurement equivalence tests should include corrections for RSs, but researchers should control for as many RSs as possible at the same time. Currently, the procedures that give the widest coverage of RS are the RIRSMACS procedure (which accounts for ARS, DARS, ERS, and MRS; Weijters et al., 2008) and the LCFA procedure (which accounts for ARS and ERS; Kieruj & Moors, 2010, 2013). The RIRSMACS procedure assumes that rating scale data are at the interval level, whereas the LCFA approach regards the data as categorical (ordinal). To accommodate research that does not ascribe the interval assumption to rating scale data but wants to cover RS, the LCFA method may need to be extended, or some other alternative to the RIRSMACS procedure may need to be developed. Perhaps this alternative will exhibit greater convergent validity with the method of (Kieruj & Moors, 2010, 2013).

According to our review of the RSs literature, although researchers have already devoted considerable attention to this topic, much still needs to be learned. We hope that we have inspired researchers to continue to expand on the boundaries of knowledge on RSs.

References

Albaum, G., Roster, C., H, Y. J., & Rogers, R. D. (2007). Simple rating scale formats: Exploring extreme response. *International Journal of Market Research*, *46*, 633–650.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, *66*(3), 374–392.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40*(6), 1235–1245.

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*(2), 229–247.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491–509.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European journal of public health*, *20*(5), 543–548.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: a two-dimensional interpretation of acquiescence. *Psychological bulletin*, *76*(3), 186–204.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research*, *36*(4), 542–562.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, *71*(5), 814–833.

Brengelmann, J. C. (1960). Extreme response set, drive level and abnormality in questionnaire rigidity. *The British Journal of Psychiatry*, *106*(442), 171–186.

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures* (Unpublished doctoral dissertation). Ghent University.

Chen, C., Lee, S.-y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, *6*(3), 170–175.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality*, *15*(1), 137–152.

Cleary, P. D., Mechanic, D., & Weiss, N. (1981). The effecit of interviewer characteristics on responses to a mental health interview. *Journal of Health and Social Behavior*, *22*(2), 183–193.

Couch, a., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of abnormal and social psychology*, *60*(2), 151–74.

Crandall, J. E. (1973). Sex differences in extreme response style: Differences in frequency of use of extreme positive and negative ratings. *The Journal* *of Social Psychology*, *89*(2), 281–293.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761–775.

De Jong, M. D., Steenkamp, J.-B. E., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115.

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, *16*(1), 20–30.

Gibbons, J. L., Zellner, J. a., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by spanish-english bilinguals. *Cross-Cultural Research*, *33*(4), 369–381.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Grimm, S. D., & Church, A. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, *33*(4), 415–441.

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192–203.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, *21*(1), 111–121.

Hofstede, G. (2001). *Cultures consequences: Comparing values, behaviors, institutions and organizations across nations* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research*, *38*(1), 52–88.

House, R. J., Hanges, P. W., Javidan, S. A., Dorfman, M., & Gupta, V. (2004). *Culture, leadership, and organizations. the globe study of 62 societies.* Thousand Oaks, CA: Sage.

Hox, J. J., de Leeuw, E. D., & Kreft, I. G. G. (1991). The effect of interviewer and resondent characteristics on the quality of survey data. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 439–461). New York: Wiley.

Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology*, *132*(2), 201–210.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*(2), 264–277.

Jordan, L. A., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, *44*(2), 210–222.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*(3), 320–342.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Kiesler, S., & Sproull, L. S. (1986). Response effects in electronic surveys. *Public Opinion Quarterly*, *50*, 402–413.

Knowles, E. S., & Condon, C. a. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379–386.

Krosnick, J. a. (1999). Survey research. *Annual review of psychology*, *50*, 537–567.

Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research*, *23*(3), 361–373.

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of personality and social psychology*, *2*(6), 907–9.

Lipps, O., & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods*, *23*(2), 156–172.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, *23*(4), 498–509.

Matarazzo, J. D., & Herman, D. O. (1984). Relationship of education and IQ in the WAIS-R standardization sample. *Journal of Consulting and Clinical Psychology*, *52*(4), 631–634.

Maxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research.

In N. Schwarz & S. Sudman (Eds.), *Context effects in social psychology research* (pp. 279–296). New York: Springer.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*(7), 1539–1550.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*(6), 779–794.

Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, *22*(1), 93–119.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of personality*, *77*(1), 261–86.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, *75*, 99–114.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*(2), 230–243.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*(3), 129–156.

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, *25*(2), 189–197.

Rowland, M. L., Naidoo, S., Abdulkadir, R., Moraru, R., Haung, B., & Pau, A. (2010). Perceptions of intimidation and bullying in dental schools: A multi-national study. *International Dental Journal*, *60*, 106–112.

Saris, W. E., Satorra, A., & Coenders, G. (2004a). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311–347.

Saris, W. E., Satorra, A., & Coenders, G. (2004b). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311–347.

Smith, T. W. (2011). Refining the total error perspective. *International Journal of Public Opinion Research*, *23*, 464–484.

Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *The Journal of Social Psychology*, *122*(2), 151–156.

Stukovsky, R., Palat, M., & Seldlakova, A. (1982). Scoring position styles in the elderly. *Studia Psychologica*, *24*, 145–154.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*(1), 91–112.

Usunier, J.-C. (2011). Language as a resource to assess cross-cultural equivalence in quantitative management research. *Journal of World Business*, *46*(3), 314–319.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Van Rosmalen, J., van Herk, H., & Groenen, P. J. (2010). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research*, *47*(1), 157–172.

Watkins, D., & Cheung, S. (1995). Culture, gender, and response buis: An analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, *26*(5), 490–504.

Weijters, B. (2006). *Response styles in consumer research* (Unpublished doctoral dissertation). Ghent University.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Geuens, M., & Schillewaert, N. (2009). The proximity effect: The role of inter-item distance on reverse-item bias. *International Journal of Research in Marketing*, *26*(1), 2–12.

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*(2), 105–121.

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological methods*, *15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*(6), 956–972.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse worded items confound measures in cross cultural consumer research? the case of the material values scale. *Journal of Consumer Research*, *30*(1), 72–91.

Zuckerman, M., & Norton, J. (1961). Response set and content factors in the california F Scale and the parental attitude research instrument. *Journal of Social Psychology*, *2*, 199–210.

CHAPTER 4

RESPONSE STYLES AND THE RURAL–URBAN DIVIDE

Response Styles and the Rural-Urban Divide

Troy Devon Thomas[1,2], Koen Abts[3] and Patrick Vander Weyden[1]

[1] *Department of Political and Social Sciences, Ghent University.*

[2] *Department of Mathematics, Physics and Statistics, University of Guyana*

[3] *Centre for Sociological Research, University of Leuven*

Abstract

This article investigates the effect of the rural–urban divide on mean response styles (RSs) and their relationships with the sociodemographic characteristics of the respondents. It uses the Representative Indicator Response Style Means and Covariance Structure (RIRSMACS) method and data from Guyana — a developing country in the Caribbean. The rural–urban divide effects substantial mean RSs differentials, and it moderates both their relationships with and the explanatory power of the respondents' sociodemographic characteristics. Within-country research is therefore subject to substantial rural–urban RSs bias, and it is hence imperative that researchers control RSs in such studies. Previous research findings should also be reexamined with RSs controlled. In addition, joint modelling of culture, RSs, and their sociodemographic predictors may clarify some of the conflicting results about their effects in the cross-cultural research literature.

Keywords: response styles, rural and urban culture, survey research, Guyana

---

Published as:

Thomas, T. D., Abts, K. & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement,* 74, 97–115.

4.1 Introduction

Rating scales are a very popular way of measuring attitudes in social sciences research (Moors, 2010); however, they can introduce error variance due to response styles (RSs). An RS is the systematic tendency of a respondent to rate items on a basis that is different from what the items are designed to measure, leading to discrepancies between the ratings and the respondents' true opinion (Paulhus, 1991). RSs compromise the validity of research results and affect substantive findings (see Baumgartner & Steenkamp, 2001; Kankaraš & Moors, 2011; Moors, 2012; Van Herk, Poortinga, & Verhallen, 2004; Welkenhuysen-Gybels, Billiet, & Cambré, 2003).

Although it is known that culture affects response styles (Harzing, 2006; Johnson, Kulesa, Cho, & Shavitt, 2005; Meisenberg & Williams, 2008), the effects of rural–urban subculture differences on RSs is not well investigated. Only one study that we know of (Arce-Ferrer, 2006) addresses rural–urban effects on RSs, but it focuses on only extreme RS (ERS). Furthermore, research on effects of RSs is mostly restricted to acquiescence RS (ARS; tendency to agree) and ERS (tendency to use the scale endpoints), whereas others like disacquiescence RS (DARS; tendency to disagree) and midpoint RS (MRS; tendency to use middle category) are investigated less often (Baumgartner & Steenkamp, 2001; Van Vaerenbergh & Thomas, 2013). In addition, the different RSs are typically modelled separately.

In this article, we investigate the effects of the rural–urban divide on the average levels of RSs and on the relationships between the RSs and their sociodemographic predictors. This provides new information about the rural–urban subculture effects. We employ joint modelling of several RSs and hence provide a more rigorous assessment of the RSs and their sociodemographic predictors in a non-Western context. The presented procedure can be replicated in cross-cultural research.

4.2 Sociodemographic Predictors of RSs

Research indicates that the respondents' age, gender, and education significantly affect RSs. However, there are many between-study inconsistencies in the findings.

Age. Many studies indicate that age is positively associated with ARS (Billiet & McClendon, 2000; Greenleaf, 1992a, 1992b; Meisenberg & Williams, 2008; Weijters, Geuens, & Schillewaert, 2010), but more inconsistencies occur for ERS. For example, while Weijters, Geuens, and Schillewaert (2010) report a positive effect of age on ERS, Johnson et al. (2005) and Moors (2008) report that there is no effect of age. For DARS and MRS, respectively, Weijters, Geuens, and Schillewaert (2010) find no effect and a positive effect of age.

Gender. There are also conflicting results for the effect of gender. Weijters, Geuens, and Schillewaert (2010) and Harzing (2006) indicate that ARS is higher among females, whereas Marin, Gamba, and Marin (1992) find no consistent gender effects. For ERS, some report no gender effect (Greenleaf, 1992b; Johnson et al., 2005; Marin et al., 1992; Moors, 2008); others report that it is higher among males (Harzing, 2006; Meisenberg & Williams, 2008); and some indicate that it is higher among females (De Jong, Steenkamp, Fox, & Baumgartner, 2008; Weijters, Geuens, & Schillewaert, 2010). In addition to this,Weijters, Geuens, and Schillewaert (2010) find no gender effect on DARS and MRS, whereas Harzing (2006) finds that MRS is higher among females.

Education. In general, education is inversely related to ARS and ERS Billiet and McClendon (2000); Greenleaf (1992b); Weijters, Geuens, and Schillewaert (2010), but Meisenberg and Williams (2008) indicate that this does not hold for ARS in countries of the South and South East Asia, whereas Moors (2008) finds no effect of education on ERS. Weijters, Geuens, and Schillewaert (2010) also find that education is inversely related to MRS and not related DARS.

Ethnicity. Research results for the effect of ethnicity are more consistent than for the other sociodemographic variables. Research consistently indicates that both ARS and ERS are higher among minority groups (Ayidiya & McClendon, 1990; Bachman & O'Malley, 1984; Baron-Epel, Kaplan, Weinstein, & Green, 2010; Clarke III, 2000; Ross & Mirowsky, 1984).

Overall, the respondent variables explain small percentages (1.4% to 8.3%) of the RSs' variance whereas culture explains large proportions (59% to 74%; Van Vaerenbergh & Thomas, 2013). Culture is therefore a major determinant. Although a smaller impact is expected due to the shorter cultural distance, we argue that subculture also affects RSs and that it is necessary to control its effects in within-country research.

4.3 Subculture Matters: The Rural-Urban Divide

In this article, we focus on subculture determined by urbanity. To establish a context, we examine the rural–urban culture debate.

Wirth's (1938) theory on rural–urban culture suggests that the greater size, density, and diversity of urban populations effect more individualism and tolerance of ambiguity than in rural populations. Gans (1962) contests these claims as being overstated, but concedes that urbanism has some relevance and that this relevance would have been more pronounced at an earlier time. Furthermore, individual characteristics must first be controlled if the effect of ecology is to be determined (Gans, 1962). Fischer (1975) also highlights the need for controlling sociodemographic variables and concludes that rural–urban subculture differences are persistent.

Mixed empirical support is found for the theories, but the evidence is mostly in favor of the urbanism theory Petković (2007); Tittle (1989). For example, Tittle (1989) and Tittle and Grasmick (2001) find that population size is positively associated with anonymity, tolerance, alienation, and deviant behaviour, and negatively associated with social bonds, thereby confirming the urbanism theory. In addition, Petković (2007) finds that rural residents are more xenophobic and conservative but that the views are similar across areas on political variety and religious choices.

The urbanism theory is useful in explaining rural–urban culture differences and it provides a way of identifying rural and urban areas, based on size, density, and diversity. Two of the main variables that distinguish rural and urban areas are individualism and intolerance of ambiguity (Wirth, 1938). These variables are also important in cross-cultural research. Furthermore, individualism and intolerance of ambiguity (uncertainty avoidance) explain cross-cultural RSs variance (e.g. Harzing, 2006). We can therefore combine what is known from cross-cultural RSs research with the current study of the rural–urban divide.

4.4 Culture and Response Styles

Individualism. Based on the Hofstede's (2001) calculations, Harzing (2006) indicates that individualism is negatively associated with ARS and MRS, but not related to ERS. These results for ARS and ERS are confirmed by Johnson et al. (2005) and by Van Herk et al. (2004) for ARS, but the latter indicate that individualism is negatively related to ERS. In contrast, De Jong et al. (2008) show that individualism is positively related to ERS.

Uncertainty avoidance. Harzing (2006) indicates that uncertainty avoidance has no relationship with ARS or MRS, but that it is positively related to ARS and ERS with respect to the Globe dimensions (House, Hanges, Javidan, Dorfman, & Gupta, 2004). In addition, several studies show that uncertainty avoidance is positively related to ERS (see Baumgartner & Steenkamp, 2001).

4.5 Hypotheses

Based on the literature, ARS is generally negatively related to individualism and either has no relationship with uncertainty avoidance (Hofstede's calculations) or is positively associated with it (Globe calculations). If the urbanism theory holds, both these effects would result in lower ARS in the urban area. Our first hypothesis is therefore the following:

Hypothesis 1: ARS is lower in urban compared with rural areas.

It is more challenging to draw a conclusion about the effect of the rural–urban divide on ERS. With a Mexican sample, Arce-Ferrer (2006) finds higher ERS in the rural area. This is consistent with the general positive effect of uncertainty avoidance on ERS. However, the findings of De Jong et al. (2008) suggest that individualism pulls ERS in the opposite direction. We take the finding of Arce-Ferrer (2006) along with the consistent effect of uncertainty avoidance as some evidence, though not unequivocal, of the net effect of the divide on ERS and hypothesize that

Hypothesis 2: ERS is lower in urban compared with rural areas.

DARS is not as well investigated in cross-cultural research as ARS and ERS. However, given that ARS is expected to be lower in the urban area, there is a strong possibility that DARS is higher in the urban area. We note this as our third hypothesis:

Hypothesis 3: DARS is higher in urban areas than in rural areas.

The available evidence of the effect of culture on MRS is limited. However, the negative association between MRS and individualism together with the lack of association with uncertainty avoidance (Harzing, 2006) suggest that MRS is lower in the urban area. We advance this as our fourth hypothesis:

Hypothesis 4: MRS is lower in urban compared with rural areas.

Turning our attention to the sociodemographic variables, we note that Meisenberg and Williams (2008) show that the effects of age and education on

ARS and ERS and the effect of gender on ARS are significant in some regions of the world but not in others. The directions of some of these effects also change from one area to another. These observations in combination with the inconsistent antecedent effects suggest that culture moderates these relationships. We therefore argue for joint estimation of the effects of culture and the respondents' characteristics and test a fifth hypothesis:

Hypothesis 5: The effects of the respondents' sociodemographic characteristics on the RSs are moderated by the rural–urban divide.

Although this implies an omnibus test of the equality of the coefficients, we go further and identify the specific respondent characteristics whose effects are moderated by the rural–urban divide.

4.6 Data and Method

4.6.1 Population and Sample

This study was conducted in Guyana, which is an English speaking, developing country on the mainland of South America. The findings are applicable to the coastal inhabitants (specifically Regions 2, 3, 4, 5, 6, and 10), who account for approximately 90.5% of the total population (Bureau of Statistics, 2002). The suggestion that the urbanism theory would have been more relevant at an earlier time (Gans, 1962) indicates that the rural–urban divide is expected to become less pronounced as societies become less traditional. As a result, rural–urban differences are expected to be greater in developing countries compared with Western countries. This was not the basis for selecting Guyana, but we believe that it (and similar countries) is a good candidate for supporting a rural–urban RSs divide. The data source is the Values and Poverty Study in Guyana (VAPO Guyana) collected between April and May 2012. The VAPO Guyana was funded by the Flemish Inter-University Counsel and jointly executed by the University of Guyana and Ghent University.

The study is designed to test methodological and substantive issues. It includes RSs and provides a unique opportunity to study the effect of the rural–urban divide in a non-Western setting. The data were collected via face-to-face interviews by a survey organization, DPMC, under the supervision of the University of Guyana and Ghent University. This organization also collected data for the Americas Barometer (Latin American Public Opinion Project) in Guyana. All the interviewers participated in a training session organized by DPMC and a subsequent briefing session organized by the VAPO research team (see Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012).

The VAPO Guyana employed a two-step sampling procedure which randomly selected municipalities with probability proportional to municipality size, and respondents within the municipalities with equal probabilities. The procedure resulted in the selection of 87 clusters within 51 municipalities. A total of 1,048 individuals were interviewed at an overall response rate of 87% (American Association for Public Opinion Research, 2011, RR2, p. 44). The data are weighted through iterative proportional fitting (Vander Weyden et al., 2012).

4.6.2 Rural-Urban Distinction

The rural–urban distinction is based on the urbanism theory. Consistent with this theory, we use the population size, density, and ethnic diversity to determine the rural and urban areas.

Region 4 is clearly the largest and most densely populated area (Table 4.1). It is not matched by any other region on these two variables. It contains the capital city and the only university in the country and it is the main economic area. The density of Region 3 (27.5 per km^2) is higher than all regions except Region 4 (139 per km^2), but the difference from Region 4 is substantial. Region 3 is physically close to, but not contiguous with Region 4 since it is separated by the Demerara River. Based on size and density, Region 3 is therefore not

Table 4.1: Regional Population and Population Densities

Region	Population	Population per Square Kilometer
Region 2	49,254	8.0
Region 3	103,061	27.5
Region 4[a]	310,320	139.0
Region 5	52,428	12.5
Region 6	123,694	3.4
Region 10	41,114	2.4

a. Defined as urban in this study.
Obtained from the website of the Guyana Bureau of Statistics.

in the same category as Region 4.

Table 4.2 shows the within-region percentage of each ethnic group. With the exception of Regions 2 and 4, a single ethnic group accounts for more than 50% of the total population of each region. These two regions are therefore more diverse. However, Region 2 does not qualify as an urban area based on size and density. Noticeably, one ethnic group accounts of 65.47% of the population of Region 3. The importance of diversity to the urbanism theory suggests that Region 3 does not qualify as urban in this regard. We therefore designate Region 4 as an urban area and all other regions as rural. With this distinction, the sample sizes are 570 and 478 for rural and urban, respectively.

4.6.3 Respondent Variables

Age is a continuous variable measured in years. Gender is dichotomous: 1 = male and 0 = female. Education represents the level of schooling completed at the time of the data collection and it is coded into three levels: 1 = up to primary schooling; 2 = secondary education; and 3 = above secondary. Ethnicity is dichotomous with 1 representing the majority group — East Indians — and 0 representing the other ethnicities (Afro, Amerindians, Chinese, Mixed, Portuguese, and White).

Table 4.2: Within-Region Ethnic Composition

Ethnicity	Region 2	Region 3	Region 4[a]	Region 5	Region 6	Region 10
African/ Black	13.41	21.23	41.67	32.55	21.06	54.98
Amerindian	16.27	2.01	1.69	1.95	1.63	7.10
Chinese	0.09	0.16	0.26	0.11	0.18	0.15
East Indian	47.91	65.47	37.54	57.76	68.68	3.08
Mixed	22.06	11.02	18.38	7.63	8.37	34.48
Portuguese	0.21	0.07	0.34	0.00	0.05	0.12
White	0.04	0.03	0.09	0.00	0.04	0.05
Other	0.00	0.00	0.03	0.00	0.00	0.03

Note. All values are percentages of the regional totals. *a*. Defined as urban in this study.

Obtained from the website of the Guyana Bureau of Statistics.

4.6.4 Data Analysis

Several methods to measure and adjust for RSs are available. In the case of continuous variables, the Representative Indicator Response Style Means and Covariance Structure (RIRSMACS) method (Weijters, Schillewaert, & Geuens, 2008) is the most comprehensive because it can include several RSs simultaneously and allows modelling of the relationships between the RSs. In this study, we construct an RIRSMACS model and adopt the recommendation to use 14 items per RS indicator with three indicators per RS factor (42 items in total; Weijters et al., 2008). In the RIRSMACS model, each of ARS, DARS, ERS, and MRS are latent variables estimated in a confirmatory factor analysis framework. To obtain the RS indicators, the 42 items which are scored on 5-point rating scales are split at random into three blocks and one indicator per RS is calculated per block as follows:

$$ARS = [f(4) + 2 * f(5)]/k,$$

$$DARS = [2 * f(1) + f(2)]/k,$$

$$ERS = [f(1) + f(5)]/k$$

and

$$MRS = f(3)/k,$$

where $f(x)$ is the frequency of the response option x and $k = 14$ is the number of items per block. The error variances of the items calculated from the same block are allowed to correlate (Weijters et al., 2008).

The models are estimated with LISREL 8.8 using maximum likelihood estimation. Scalar invariance across the two groups — rural versus urban — is evaluated leading to comparisons of the factor intercepts. The sociodemographic variables are then added and their effects are evaluated in a structural equation modelling framework. We also reevaluate the differences in the factor intercepts after controlling the effects of the sociodemographic variables. Because of the large sample sizes, these models are evaluated with alternative fit indices (Chen, 2007). We use root mean square error of approximation less than 0.06, comparative fit index greater than 0.95, and standardized root mean square residual less than 0.05 as benchmarks for acceptable overall fit (Byrne, Shavelson, & Muthen, 1989; Hu & Bentler, 1999).

For the estimation of RSs, the content of the items must be controlled to avoid confounding RSs with the content. This is can be achieved by randomly selecting one item per construct throughout the questionnaire (Weijters et al., 2008). In the VAPO Guyana, 45 attitude items were selected from various constructs covering several topics (including government, politics, society, crime gender roles, and many more). These items were piloted in a PAPI survey among students (n = 1,000) at the University of Guyana leading to the selection of 35 items with low correlations. We use these 35 items to measure RSs along with 7 additional items that are selected at random from indicators of constructs included in the questionnaire (see items in Appendix A.1). The average interitem correlation of the 42 items is 0.05.

The scale format is also important when analyzing RSs. Although the number of response categories do not affect ERS with end-labelled scales (Kieruj & Moors, 2010, 2013), ERS decreases with more response categories and with fully labelled scales (Weijters, Cabooter, & Schillewaert, 2010). Scale format should therefore be consistent over all the items used to measure RSs. Scales with 5 to 7 points are recommended and this is consistent with the findings in the general literature on the optimal number of scale categories (e.g. Preston & Colman, 2000). In this study, the RSs are measured by 5-point, fully labelled rating scales, from completely disagree to completely agree. The agree/disagree format tends to increase ARS by approximately 10% (Krosnick, 1999), but the scales are uniform across the groups.

4.7 Results

The Cronbach alpha for the RSs factors (see Table 4.4) are greater than 0.70 in each group except for DARS in the rural group ($\alpha = 0.65$). The results for DARS should therefore be interpreted with caution.

Initial estimation of the four-factor RIRSMACS model fails due to a non-positive definite input matrix. To avoid biased standard errors (McQuitty, 1997), we modify the model and instead estimate two separate three-factor models: one containing ARS, ERS, and DARS and the other containing ARS, ERS, and MRS. ARS and ERS are always included because they are the most recognized RSs and because joint modelling of the RSs provides a more stringent assessment of the effects of the respondent characteristics. It is expected that the two sets of models will show small variations in the estimates for ARS and ERS, since the third factor will have some effect on the relationships.

Each of the estimated models fit adequately (Table 4.3). With the exception of the DARS factor, the standardized factor loadings (Table 4.4) are greater than 0.70 and the average variance extracted is greater than 0.50, indicating the achievement of convergent validity. In addition, the square root of the average variance extracted for each factor is larger than the correlations between pairs of factors per model indicating that discriminant validity is achieved in each case (Fornell & Larcker, 1981). Consistent with the interpretation from the alpha values, the results for convergent validity indicate a need for caution when interpreting the results for DARS.

The rural–urban comparison of the factor means are done with two separate, two-group CFA models. The models are distinguished by whether they contain DARS or MRS. Both models show configural and metric invariance (Table 4.3: Mean Comparisons; (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). However, in both cases, partial scalar invariance is achieved due to the noninvariance of an item intercept (Byrne et al., 1989). For the

Table 4.3: Fit Statistics and Indices

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	$RMSEA$	$\Delta RMSEA$	CFI	ΔCFI	$SRMR$
Measurement models									
Rural^d	44.436*	15			0.055		0.993		0.023
Urban^d	23.799	15			0.036		0.998		0.025
Rural^m	51.176*	15			0.063		0.993		0.040
Urban^m	45.726*	15			0.066		0.995		0.036
Mean comparisons									
Configural invariance^d	68.235*	30			0.048		0.996		
Metric invariance^d	72.932*	36	4.697	6	0.043	−0.005	0.996	0.000	

RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; SEM = structural equation modeling; DARS = disacquiescence response style; MRS = midpoint response style. *d*. Relevant to the DARS model. *m*. Relevant to the MRS model. The $\Delta\chi^2$ for metric and scalar invariance are evaluated Bonferroni correction for multiple planned hypotheses. For an overall 5% significance level, the critical p values are .05/6 = .008 for metric invariance and .05/12 = .004 for scalar invariance. *Indicates significance at the 5% level.

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	$RMSEA$	$\Delta RMSEA$	CFI	ΔCFI	$SRMR$
Scalar invariance 1^d	143.377*	42	75.142*	12	0.068	0.020	0.989	−0.007	
Scalar invariance 2^d	125.521*	41	57.286*	11	0.062	0.014	0.991	0.005	
Configural invarince^m	96.892*	30			0.062		0.994		
Metric invariance^m	107.817*	36	10.925	6	0.061	0.001	0.994	0.000	
Scalar invariance 1^m	175.332*	42	78.440*	12	0.078	0.016	0.988	−0.006	
Scalar invariance 2^m	136.354*	40	39.463*	10	0.067	0.005	0.992	−0.002	
Sociodemographic characteristics									
SEM baseline^d	218.303*	90			0.052		0.987		
Structural invariance^d	269.054*	102	50.751*	12	0.055	0.003	0.983	−0.004	
SEM baseline^m	157.394*	90			0.059		0.986		

RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; SEM = structural equation modeling; DARS = disacquiescence response style; MRS = midpoint response style. *d*. Relevant to the DARS model. *m*. Relevant to the MRS model. The $\Delta\chi^2$ for metric and scalar invariance are evaluated Bonferroni correction for multiple planned hypotheses. For an overall 5% significance level, the critical p values are .05/6 = .008 for metric invariance and .05/12 = .004 for scalar invariance. *Indicates significance at the 5% level.

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*RMSEA*	Δ*RMSEA*	*CFI*	Δ*CFI*	*SRMR*
Structural invariance^m	306.934*	102	49.540*	12	0.061	0.002	0.983	−0.003	

RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; SEM = structural equation modeling; DARS = disacquiescence response style; MRS = midpoint response style. *d*. Relevant to the DARS model. *m*. Relevant to the MRS model. The $\Delta\chi^2$ for metric and scalar invariance are evaluated Bonferroni correction for multiple planned hypotheses. For an overall 5% significance level, the critical p values are .05/6 = .008 for metric invariance and .05/12 = .004 for scalar invariance. *Indicates significance at the 5% level.

Table 4.4: Standardised Factor Loadings

		Rural				Urban			
		ARS	DARS	ERS	MRS	ARS	DARS	ERS	MRS
DARS model	ARS1	0.79				0.90			
	ARS2	0.80				0.92			
	ARS3	0.72				0.83			
	DARS1		0.65				0.72		
	DARS2		0.58				0.66		
	DARS3		0.57				0.64		
	ERS1			0.87				0.97	
	ERS2			0.94				0.97	
	ERS3			0.90				0.94	
MRS model	ARS1	0.78				0.89			
	ARS2	0.79				0.92			
	ARS3	0.75				0.84			
	ERS1			0.86				0.97	
	ERS2			0.92				0.97	
	ERS3			0.90				0.93	
	MRS1				0.94				0.84
	MRS2				0.81				0.84
	MRS3				0.73				0.91
AVE[d]		0.60	0.36	0.82		0.78	0.46	0.91	
AVE[m]		0.60		0.80	0.63	0.78		0.91	0.75
Alpha[a]		0.81	0.62	0.92	0.83	0.91	0.71	0.97	0.86

ARS = acquiescence response style; DARS = disacquiescence response style; ERS = extreme response style; MRS = midpoint response style; AVE = average variance extracted. *a*. Relevant to both models. *d*. Relevant to the DARS model. *m*. Relevant to the MRS model.

Table 4.5: Factor Mean Difference

Model	Response style	Difference	SE	t Value	Effect	Effect†
DARS model	ARS	−0.04	0.02	−2.41*	0.22*	0.24*
	DARS	0.09	0.01	9.77*	0.90*	0.80*
	ERS	0.07	0.01	4.87*	0.44*	0.35*
MRS model	ARS	−0.05	0.02	−3.11*	0.29*	0.35*
	ERS	0.06	0.02	3.83*	0.38*	0.29*
	MRS	0.04	0.01	3.79*	0.33*	0.29*

ARS = acquiescence response style; DARS = disacquiescence response style; ERS = extreme response style; MRS = midpoint response style. The effect sizes are for the factor mean differences with rural as the reference group. The effect sizes are scaled by the standard deviations of the factors in the rural group. Difference = mean difference. Effect†= effect with sociodemographics controlled.
*Indicates significance at the 5% level.

model with DARS, the freed item intercept is that of the first ARS indicator (modification index 17.27), whereas for the model with MRS the intercept of the second ERS indicator (modification index 37.848) is freed. As a result, the identified items do not contribute to the mean difference in the respective factors in the respective models. We also highlight that $\Delta\chi^2$ for the partial scalar invariance models remain significant, but that $\Delta RMSEA$ and ΔCFI are negligible (see Chen, 2007).

In both models, the factor mean for each RS is significantly different between the rural and urban groups even after controlling the effects of the sociodemographic variables (Table 4.5). The directions of the mean differences of ARS and ERS are also consistent, but the effect sizes show small between-model variations as expected. Overall, urban residents agree less often, but they are more likely to give extreme responses, disagree, and use the scale midpoint than rural residents. These results confirm the first and third hypotheses, but falsify the second and fourth hypotheses.

Despite this, we can conclude that the rural–urban divide results in significant mean differences in RSs beyond that explained by the respondent characteristics and that it potentially introduces substantial ecological bias into within-country research results. The results from the structural equation models with between-group equality constraints on the factor loadings (Table 4.3: Sociodemographic characteristics) indicate that the sociodemographic variables explain larger percentages of the variance in the RSs in the rural group. The explained variances are within the range indicated by the literature except for ARS and DARS in the rural group where they are marginally higher (Table 4.6). However, the between-group differences in explained variances are large. The explanatory powers in the rural group are approximately 2.08, 1.50, 1.74, and 5.64 times higher for ARS, ERS, DARS, and MRS, respectively, than in the urban group. The rural–urban divide therefore moderates the explanatory powers of the sociodemographic predictors of RSs.

Tests for the between-group equality of the effects (Table 4.3: Regression invariance) return significant $\Delta\chi^2$ statistics, thus confirming that the rural–urban divide moderates the effects of the respondents' characteristics on the RSs. Specifically, the effects of education and ethnicity are susceptible to rural–urban moderation. In the model containing DARS, significant modification indices occur for the effects of education on ARS (20.35), ERS (20.71), and DARS (18.50) and for the effects of ethnicity on ARS (19.53) and ERS (9.26). In the model containing MRS, significant modification indices are observed for the effects of education on MRS (17.98) and for the effects of ethnicity on ARS (8.25) and ERS (6.84). The moderating effect of the rural–urban divide is consistent for the impact of ethnicity but not for the impact of education on ARS and ERS.

Examination of the within-group standardized effects (Table 4.6) reveals that with the exception of the expected small between-model fluctuations, the results are consistent for ARS and ERS. The results lead to a few conclusions about specific variables. First, age is negatively related to ARS in the urban area, whereas ERS increases with age in the rural area. Second, gender predicts only ARS in the rural group. In the rural area, males are less likely than females

Table 4.6: Within-Group Standardized Structural Relationships

Predictor	Urban				Rural			
	ARS	DARS	ERS	MRS	ARS	DARS	ERS	MRS
Age^d	−0.11*	0.02	−0.08		0.05	0.09	0.10*	
Male^d	−0.05	0.08	−0.01		−0.10*	0.07	−0.04	
Education^d	−0.20*	0.18*	−0.13*		−0.25*	0.06	−0.07	
Majority group^d	−0.16*	−0.08	−0.17*		0.04	−0.27*	−0.18*	
Age^m	−0.11*		−0.08	0.07	0.05		0.10*	−0.01
Male^m	−0.04		−0.01	0.01	−0.09*		−0.04	0.02
Education^m	−0.20*		−0.13*	−0.01	−0.25*		−0.07	0.23*
Majority group^m	−0.16*		−0.17*	0.09	0.04		−0.17*	−0.10*
R-squared^d	4.7%	5.3%	3.1%		9.7%	9.2%	4.7%	
R-squared^m	4.6%		3.1%	1.4%	9.7%		4.6%	7.9%

Note. ARS = acquiescence response style; DARS = disacquiescence response style; ERS = extreme response style; MRS = midpoint response style. d. Relevant to the DARS model. m. Relevant to the MRS model.
*Indicates significance at the 5% level.

to use ARS. Third, more education lowers ARS overall and lowers ERS in the urban area, but increases DARS among urban residents and increases MRS among rural residents. Fourth, the majority ethnic group — East Indian — uses less ERS overall, less ARS in the urban area, and less DARS and MRS in the rural area. Consistent with the moderating role of the rural–urban divide, these results indicate that whether or not the antecedent effects are significant depend on the group considered.

4.8 Discussion

This study finds mean RSs differentials between rural and urban areas and confirms that the rural–urban divide moderates the effects of the respondent characteristics on the RSs. Specifically, the divide moderates the effects of education and ethnicity on some of the RSs. Within-country research results are therefore subject to RSs bias. The immediate question that arises is about whether or not the bias is substantial enough to have high impact. In their study, Weijters et al. (2008) find the largest effect size between data collection modes to be 0.47 and the smallest to be 0.18. They also show that both the factor loadings and the factor means of a substantive construct are substantially biased when the RSs are not controlled. We find that all the effect sizes are larger than 0.18 and that the effect size for DARS is larger than 0.47 after controlling the sociodemographic variables. As a result, not only data collection methods but also the rural–urban divide can lead to substantial differential RSs bias — at least in a non-Western context. Consequently, within-country research results may accurately represent neither the country as a whole nor subgroups within the country. It is therefore imperative that researchers explicitly control the rural–urban RSs bias in within-country research.

Higher ERS in the urban area is in conflict with the finding in Mexico (Arce-Ferrer, 2006). Given that intolerance of ambiguity and individualism may pull ERS in different directions, it seems that individualism has a higher impact than intolerance of ambiguity on ERS in Guyana compared with Mexico. Because of this, a priori determination of the net rural–urban effect on ERS in a new context may continue to be difficult. However, the rural–urban divide does affect ERS.

The explanatory powers for ARS, ERS, DARS, and MRS in the urban area and for ERS and MRS in the rural area (1.4% to 7.9%) are generally consistent with the literature (e.g. De Jong et al., 2008; Meisenberg & Williams, 2008; Weijters, Geuens, & Schillewaert, 2010), but the explained variance for ARS and DARS (9.2% to 9.7%) are marginally higher than expected. Overall, the explanatory powers tend to be higher in the rural area. With respect to the pattern of significant effects, each of age, gender, education, and ethnicity are susceptible to cultural moderation and this may explain some of the inconsistencies in the antecedent relationships in cross-cultural research. As a result, the effects of respondents' sociodemographic characteristics on RSs should be reevaluated in cross-cultural research to determine the moderating role of culture.

With respect to the impact of the respondents' characteristics across the groups, we find that the effects of age and gender are equal whereas the effects of education on MRS and of ethnicity on ARS and ERS are moderated by the rural–urban divide. The effect of the rural–urban divide on the impact of education on ARS and MRS also seem to depend on whether MRS or DARS is modelled jointly with ARS and ERS. This highlights the importance of modelling the RSs jointly to provide more stringent evaluations.

The falsification of our hypotheses about the mean differences in ERS and MRS should not be interpreted as limited support for the urbanism theory. Although the net effect on ERS seems to depend on the relative impact of individualism and intolerance of ambiguity, studies including MRS are limited. There is more consistent evidence for ARS and in this case the hypothesis is confirmed. We emphasize, however, that the validity of Wirth's (1938) causal inferences is still to be clarified by sociologists (Tittle & Grasmick, 2001). Furthermore, individualism and tolerance of ambiguity are not measured in this study. Their relative levels in the regions of Guyana are inferred from the urbanism theory. Gans's (1962) claim leads us to believe that the theory applies to the more traditional Guyanese culture, but it is possible that the effects detected are due to other variables. For example, Arce-Ferrer (2006) attributed the difference in ERS between rural and urban schools to differences in familiarity with rating scales. Despite the growing popularity of nationwide surveys in Guyana, rural–urban differences in familiarity with rating scales may also exist and may explain the differences in the levels of RSs. However, our main positions are that the rural–urban divide with respect to RSs is substantive in Guyana and we expect that it also exists other non-Western societies. Therefore, more attention should be paid to the differential rural–urban RSs bias which can affect research results.

4.9 Conclusion

The rural–urban divide influences differential levels of RSs and these rural–urban differentials potentially affect all survey research findings whenever rating scales are used. Urban residents tend to agree less, but give extreme responses, disagree, and use the scale midpoint more often than rural residents. These differences cannot be explained by respondents' sociodemographic characteristics; hence, it is necessary for researchers to explicitly control RSs in within-country research. Our study focuses on a developing country. The more traditional culture may explain the existence of a rural–urban RSs divide, but the divide should also be investigated in Western countries. The existence of within-country RSs differentials calls into question the validity of all research results for developing countries in which rating scales are used, but in which rural–urban differences in RSs are not controlled. Substantive theories formulated and supported by survey data should therefore be reexamined. The rural–urban divide also moderates the effects of education and ethnicity on some RSs and it results in differences in the significant and non-significant effects of the sociodemographic variables. Extrapolating this to the cross-cultural setting, we argue for joint modelling of the effects of culture and the respondents' sociodemographic characteristics on the RSs in combination with the simultaneous modelling of several RSs to clarify some of the inconsistencies in the literature.

References

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Deerfield, IL: AAPOR.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, *66*(3), 374–392.

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*(2), 229–247.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491–509.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European journal of public health*, *20*(5), 543–548.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Bureau of Statistics. (2002). 2002 population & housing census. Retrieved from `http://www.statisticsguyana.gov.gy/census.html`

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement in variance. *Psychological Bulletin*(3), 456–466.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality*, *15*(1), 137–152.

De Jong, M. D., Steenkamp, J.-B. E., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115.

Fischer, C. S. (1975). Toward a subcultural theory of urbanism. *American Journal of Sociology*, *80*(6), 1319–1341.

Fornell, C., & Larcker, D. F. (1981). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Journal of Marketing Research*, *18*, 39–50.

Gans, H. (1962). Urbanism and suburbanism as ways of life: A re-evaluation of definitions. In A. M. Rose (Ed.), *Human behavior and social processes; an interactionist approach* (pp. 625–648). Boston: Houghton-Miffl.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Hofstede, G. (2001). *Cultures consequences: Comparing values, behaviors, institutions and organizations across nations* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

House, R. J., Hanges, P. W., Javidan, S. A., Dorfman, M., & Gupta, V. (2004).

Culture, leadership, and organizations. the globe study of 62 societies. Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*(2), 264–277.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*(3), 320–342.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Krosnick, J. a. (1999). Survey research. *Annual review of psychology*, *50*, 537–567.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, *23*(4), 498–509.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*(7), 1539–1550.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*(6), 779–794.

Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, *22*(1), 93–119.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Petković, J. (2007). Traditional values and modernization challenges in forming urban and rural culture. *Philosophy, Sociology and Psychology*, *6*, 23–39.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, *25*(2), 189–197.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.

Tittle, C. R. (1989). Influences on urbanism: A test of predictions from three perspectives. *Social Problems*, *36*(3), 270–288.

Tittle, C. R., & Grasmick, H. G. (2001). Urbanity: Influences of urbanness, structure, and culture. *Social Science Research*, *30*(2), 313–335.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana.* Ghent, Belgium: Ghent University.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological methods*, *15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, *44*(1), 1–24.

CHAPTER 5

MEASUREMENT INVARIANCE, RESPONSE STYLES AND RURAL–URBAN MEASUREMENT COMPARABILITY

Measurement Invariance, Response Styles and Rural-Urban Measurement Comparability

Troy Devon Thomas[1,2], Koen Abts[3] and Patrick Vander Weyden[1,4]

[1] *Department of Political and Social Sciences, Ghent University.*

[2] *Department of Mathematics, Physics and Statistics, University of Guyana*

[3] *Centre for Sociological Research, University of Leuven*

[4] *Founder of the independent research company — FocusUP.*

Abstract

This paper investigates the effect of response styles (RSs) on rural-urban measurement comparability in Guyana. It uses the representative indicators response styles means and covariance structure (RIRSMACS) model and finds that traditional measurement invariance (MI) tests provide inadequate assurance of the absence of rural–urban measurement bias when RSs are not controlled. Even when MI is achieved, RSs can still differentially affect measurements and substantive results between rural and urban regions. In addition, a lack of MI may be at least partially due to RSs bias, but MI may also be due to RSs. Therefore, adjustments for RSs are necessary and researchers should be cautious about pooling data across rural and urban areas without controlling RSs.

Keywords: response styles, measurement invariance, rural and urban culture, item bias

---

Accepted for publication at the Journal of Cross-Cultural Psychology

5.1 Introduction

Response styles (RSs) are the respondents' systematic tendencies to respond to rating scale items in certain ways regardless of the content (Baumgartner & Steenkamp, 2001). They are linked to culture and are known to bias the results of cross-cultural research (Johnson, Kulesa, Cho, & Shavitt, 2005; Locke & Baik, 2009; P. B. Smith, 2004; T. W. Smith, 2011; Van Herk, Poortinga, & Verhallen, 2004; Van Vaerenbergh & Thomas, 2013). However, although rural and urban areas may have different cultures (Wirth, 1938), the impacts of RSs on measurements and research results between such areas have not been investigated. Besides, measurement invariance (MI) is required for cross-cultural (and group) comparisons (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000; F. J. R. Van de Vijver & Tanzer, 2004), but traditional MI tests do not necessarily detect RSs bias (Weijters, Schillewaert, & Geuens, 2008).

This paper investigates the effects of RSs on the comparisons of measurements between rural and urban areas. It provides a first impression of the within-country, differential effects of RSs on measurements and research results and examines the effectiveness of traditional MI tests in detecting RSs bias. These objectives are achieved by evaluating MI between rural and urban areas with and without adjustments for RSs.

5.2 Measurement Comparability

5.2.1 Bias

In multi-group research, valid group comparisons assume the absence of bias (Kankaraš & Moors, 2010). There are three main types of bias: (1) construct bias occurs when the construct measured is not the same across the groups; (2) method bias refers to cultural factors affecting most or all the items in an instrument; and (3) item bias results from idiosyncrasies of specific items in the questionnaire. Controlling bias is especially important when studying cultural groups since larger cultural distances increase the chances and the size of bias (F. Van de Vijver & Leung, 1997; F. J. R. Van de Vijver & Poortinga, 1997; F. J. R. Van de Vijver & Tanzer, 2004). If cultural differences coincide with the rural–urban divide, particular attention to bias is warranted when dealing with data collected across rural and urban areas.

5.2.2 Rural–Urban Culture and RSs

Wirth (1938) indicates that individualism and tolerance of ambiguity are higher in urban compared to rural areas and that these cultural differences result from the greater size, density and diversity of urban populations. Although Gans (1962) suggests that these claims would only have been relevant at an earlier time, C. S. Fischer (1975) indicates that the rural–urban subcultural differences are still persistent. Overall, the empirical evidence is mostly in support of Wirth's urbanism theory (Petković, 2007; Tittle, 1989). For example, Tittle and Grasmick (2001) find that population size is positively associated with anonymity, tolerance, alienation, and deviant behaviour, but negatively associated with social bonds. The urbanism theory therefore provides useful characterisations of rural and urban cultures.

Given the cultural differences, mean RSs differentials are likely to exist between rural and urban groups. This is the case in Guyana where acquiescence RS (ARS: tendency to agree) is lower in the urban region and extreme RS (ERS: tendency to use scale endpoints), disacquiescence RS (DARS: tendency to disagree) and midpoint RS (MRS: tendency to use scale midpoint) are higher in the urban region (Thomas, Abts, & Vander Weyden, 2014). In addition, ERS is lower in the urban area in Mexico (Arce-Ferrer, 2006). Although within-country, rural–urban cultural distance is expected to be smaller than that between the groups commonly found in cross-cultural research (countries), the absence of bias must be demonstrated and not assumed (F. J. R. Van de Vijver & Poortinga, 1997). At the data analysis stage, measurement comparability is determined from MI evaluations.

5.3 MI

MI implies independence between observed scores and group membership given the true score on a construct (Meredith, 1993; Meredith & Millsap, 1992; Millsap, 1995). When MI is achieved, members of different groups with the same position on the construct of interest are expected to have the same observed score (Millsap, 1997; Schmitt & Kuljanin, 2008). A lack of MI invalidates group comparisons (Byrne & Watkins, 2003; Chen, 2008; Cheung & Rensvold, 1999; Kankaraš & Moors, 2010; Oort, Visser, & Sprangers, 2009; Van der Veld & Saris, 2011a; F. J. R. Van de Vijver & Tanzer, 2004). However, there are different levels of MI and each level permits a different kind of comparison. The levels of MI that are of primary interest in cross-cultural research are configural, metric and scalar invariance (F. Van de Vijver & Leung, 1997).

Configural Invariance. Configural invariance means that the number of factors is the same and that the models have a fixed pattern of salient and non-salient factor loadings across the groups (Horn & McCardle, 1992; Steenkamp & Baumgartner, 1998). As such, the same constructs are measured in each group. This is a basic requirement for all subsequent levels of MI and it is affected by construct bias (F. Van de Vijver & Leung, 1997).

Metric Invariance. Metric invariance asserts that the factor loadings are equal across the groups (Dimitrov, 2010; Steinmetz, Schmidt, Tina-Booh, Wieczorek, & Schwartz, 2009). The interpretations of the items are therefore preserved and this permits comparisons of structural relationships (Dimitrov, 2010). Both method and item bias affect metric invariance (F. Van de Vijver & Leung, 1997).

Scalar Invariance. Scalar invariance indicates that the item intercepts are equal across the groups (Dimitrov, 2010; Sass, 2011; Schmitt & Kuljanin, 2008; Steinmetz et al., 2009). Scalar invariance requires metric invariance. It permits comparisons of factor means and is affected by construct, method and item bias (F. Van de Vijver & Leung, 1997).

Partial Invariance. More restrictive levels of MI are less likely to be achieved (Horn & McCardle, 1992; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). Nevertheless, ignoring a lack of MI is dangerous since the bias can affect research results (Chen, 2008; Millsap, 2007, 2010; Millsap & Yun-tein, 2004; Sass, 2011). If bias is detected in some items, comparisons are still possible under partial MI which excludes the affected items from the comparisons (Byrne, Shavelson, & Muthen, 1989). However, partial MI can lead to further problems. Modification of the measurement model to permit comparisons may result in capitalising on chance. Furthermore, the selected referent item may affect whether or not non-invariant items are identified and which items are non-invariant (Cheung & Rensvold, 1999). When several items are non-invariant or where the bias caused is large, partial MI can result in substantial changes in the meaning of the construct and it may also influence substantive research outcomes (Millsap & Yun-tein, 2004). While partial invariance is an alternative to abandoning comparisons altogether when biased items are encountered, it should be used with caution.

Corrections for RSs. Traditional MI tests do not necessarily detect RSs bias, yet RSs can affect metric and scalar invariance. RSs can affect factor loadings and group locations on constructs and thus distort research results (Kankaraš & Moors, 2011; Weijters et al., 2008; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). In particular, higher(lower) ERS can increase(decrease) factor loadings whereas higher(lower) ARS increases(decreases) item means (Cheung & Rensvold, 2000). RSs may also inflate(reduce) scale variances which affects whether or not significant differences are detected (Baumgartner & Steenkamp, 2001). If RSs vary systematically between the groups under study, they may either artificially result in or hinder MI; both of which are undesirable (Weijters et al., 2008).

Given that the mean levels of RSs differ between rural and urban areas, we believe that rural–urban measurement comparability could be affected. We expect that RSs affect factor convergent validity and MI assessments and bias comparisons of factor means.

5.4 Data and Methods

5.4.1 Data

The data used in this study are obtained from the Values and Poverty Study in Guyana (VAPO Guyana) which was conducted between April and May 2012. The study was funded by the Flemish Inter-University Counsel (VLIR) and jointly executed by the University of Guyana and Ghent University. It investigates both methodological and substantive issues and provides an opportunity to study the effect of the rural–urban RSs divide in a non-Western setting. The data were collected via face-to-face interviews by a survey organisation (DPMC) under the supervision of the Universities of Guyana and Ghent. This organisation also collected data for the Americas Barometer (Latin American Public Opinion Project) in Guyana. The interviewers who participated in the study were trained by DPMC and they attended a two-day briefing session organised by the VAPO research team (Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012).

The VAPO Guyana employed a two-step sampling procedure which randomly selected municipalities with probability proportional to size, and respondents within the municipalities with equal probabilities. This procedure resulted in the selection of 87 clusters within 51 municipalities. In total, 1048 individuals were interviewed at a response rate of 87% (American Association for Public Opinion Research, 2011, RR2). The data are weighted by iterative proportional fitting.

5.4.2 Variables and Measures

Rural–urban Distinction. The rural–urban distinction is based on the urbanism theory. Population size, density and ethnic diversity are used to determine the rural and urban areas (Wirth, 1938). Region 4 is clearly the largest and most densely populated (Table 5.1). It contains the capital city and it is the main economic area. The population density of Region 3 is larger than all except Region 4, but it is not at the same level as Region 4. Although the population of Region 6 is quite large, it is scattered over a large area resulting is a very low density. This region is therefore not urban. Region 2 and 4 are the only areas for which a single ethnic group does not account for more than 50% of the total population (Table 5.1) and are hence more diverse. However, Region 2 is not urban with respect to size and density. Region 4 is therefore regarded as urban and all others as rural. This distinction results in sample sizes of 570 and 478 for the rural and the urban group respectively.

Attitude Constructs. Four attitude constructs are used in this study: political cynicism, perceived discrimination, economic uncertainty and social (dis)trust. First, political cynicism is a generalized negative attitude of suspiciousness about and disdain for the motives, sincerity and conduct of politicians and politics. Second, perceived discrimination measures feelings of relative deprivation emanating from perceived unequal treatment and relative shortcomings compared to others in regard to public policy resulting in feelings of social injustice. Third, economic insecurity refers to increased feelings of vulnerability at the labour market and negative expectations about one's future socio-economic position. Fourth, social (dis)trust refers to the unwillingness to be vulnerable in situations of risk and dependency reflecting a lack of belief in the sincerity and good intentions of others (Abts, 2012). The items (see Table 5.2) measuring these constructs are validated in other surveys like the General Election Studies in Belgium (see Abts, 2012; Swyngedouw, Abts, & Rink, 2009). Each item is scored on a 5-point fully labelled rating scale (Dis-

Table 5.1: Regional Population Size, Density and Diversity

Region	Population	Density	Ethnicity				
			African/ Black	Amerindian	East Indian	Mixed	Other
Region 2	49,254	8.0	13.41	16.27	47.91	22.06	0.34
Region 3	103,061	27.5	21.23	2.01	65.47	11.02	0.26
Region 4	310,320	139.0	41.67	1.69	37.54	18.38	0.72
Region 5	52,428	12.5	32.55	1.95	57.76	7.63	0.11
Region 6	123,694	3.4	21.06	1.63	68.68	8.37	0.27
Region 10	41,114	2.4	54.98	7.10	3.08	34.48	0.35

Other consists of Chinese, Portuguese, White and other ethnicities. Note: Values are stated as percentages of the regional totals. Density is the population per square kilometre.
Obtained from Bureau of Statistics (2002).

agree/Agree) and the constructs have adequate Cronbach's alpha reliability in both the rural and urban groups (Table 5.2) and the factor loadings are adequate (Table 5.4 and Table 5.7).

5.4.3 Methods

We estimate the RSs with the representative indicators response styles means and covariance (RIRSMACS) model which uses a confirmatory factor analysis framework and which regards the measurements as continuous. The RIRSMACS model includes ARS, ERS, DARS and MRS as latent variables each having three indicators calculated from three blocks of items (one indicator each per block) (Weijters et al., 2008).

For estimating RSs, the content of the items must be controlled to avoid confounding content with style. In the VAPO Guyana, 45 attitude items were randomly selected from various constructs covering several topics (including government, politics, society, crime gender roles and many more). These items were tested in a PAPI survey among students (n=1000) at the University of Guyana leading to the selection of 35 items with low inter-correlations ($|r| \leq 0.30$). The selected items were then included in the larger VAPO Guyana questionnaire to ensure that separate items are always available to measure RSs in addition to the substantive constructs included (Vander Weyden et al., 2012).[1] In this study, the RSs are measured by a random selection of 27 of these items (See Appendix A.2) and they have an average interitem correlation of 0.06. This number of items is larger than the recommended total of at least 15 for corrections with the RIRSMACS model (Weijters et al., 2008). To control the impact of scale format, the RSs items are all scored on 5-point fully labelled rating scales that are identical to those of the content items except for economic insecurity which has different verbal labels.

[1]Neither pretesting nor separate items are absolutely necessary for controlling RSs. Usually, heterogeneous item can be obtained by randomly selecting one item per construct from the constructs that are included the questionnaire provided that the questionnaire covers a variety of topic areas.

Table 5.2: Constructs and Items

Construct	Item Code	Item	Cronbach's Alpha	
			Rural	Urban
Political Cynicism	CYN1	It makes no sense to vote; the parties do what they want to do anyway.	0.75	0.78
	CYN2	Parties are only interested in my vote, not in my opinion.		
	CYN3	Most politicians promise a lot, but don't do anything.		
	CYN4	All politicians are profiteers.		
Perceived Discrimination	DISC1	If we need something from the government, people like me have to wait longer than others.	0.74	0.90
	DISC2	People like me are being systematically neglected, whereas other groups received more than they deserve.		
	DISC3	The government does a lot more for other ethnic groups than for us.		
Economic Insecurity	INSE1	How much are you worried that your financial worries will increase in the coming years?	0.81	0.86
	INSE2	How much are you worried that you will have difficulties in keeping your financial position?		
	INSE2	How much are you worried that your children and the coming generation will have it much more difficult?		
Social (Dis)Trust	DIST1	These days, you really don't know who you can trust.	0.67	0.83
	DIST2	Today you cannot be careful enough when dealing with other people.		

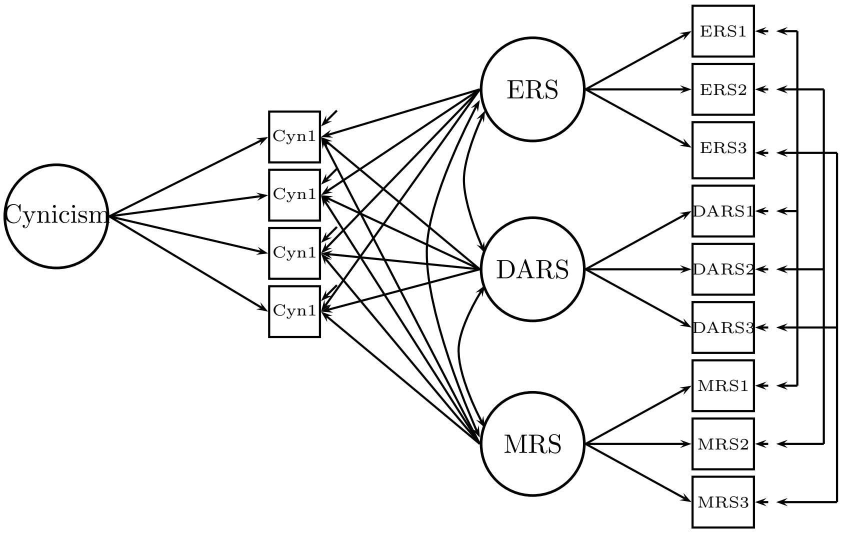

Figure 5.1: Correcting for RSs using the RIRSMACS model

To obtain the values of the RSs indicators, the pool of items is divided at random into three blocks of 9 items each and one indicator per RS is calculated from each block. These values are calculated as:

$$ARS = [f(4) + 2 * f(5)]/k,$$

$$DARS = [2 * f(1) + f(2)]/k,$$

$$ERS = [f(1) + f(5)]/k$$

and

$$MRS = f(3)/k,$$

where $f(x)$ is the frequency of the response option x and k is the number of items per block (Weijters et al., 2008).

In the RIRSMACS model, each indicator of the attitude constructs is also modelled as an indicator of each RS and the RSs are not allowed to correlate with the attitude constructs (see Figure 5.1). The impacts of a single RS on the items measuring a particular attitude construct are held equal, but the effects of different RSs on the same set of items are allowed to be different. These basic constraints are extended to the model which includes more than one attitude construct by simply regarding each construct as a different unit and applying the basic constraints. In this case, the impacts of an RS on sets of items measuring different constructs, are not equated since the same RS can differentially affect different constructs. Between-group equality constraints on these effects are also not imposed since the RSs may affect the items differentially between the groups. Finally, covariances between the error terms of the RSs indicators that are calculated from the same block of items are estimated freely but all others are fixed to zero (Weijters et al., 2008).

The use of the RIRSMACS method presupposes multi-group confirmatory factor analysis to evaluate MI. This method is the most popular and the most powerful and versatile for testing MI (Meuleman, Davidov, & Billiet, 2009; Steenkamp & Baumgartner, 1998). The models are estimated using the robust maximum likelihood estimator in Mplus 7.11 and evaluated with a combination of fit indices and by using Jrule for Mplus (Oberski, 2008; Van der Veld, 2008).[2] MI is evaluated twice for each set of attitude constructs; once without and once with corrections for RSs.

[2]Due to the large sample sizes, the models are evaluated with alternative fit indices (Chen, 2007). We use the Root Mean Square Error of Approximation (RMSEA) less than or equal to 0.06, the Comparative Fit Index (CFI) greater than or equal to 0.95 and the Standardised Root Mean Square Residual (SRMR) less than or equal to 0.05 to indicate acceptable overall fit (Byrne et al., 1989; Hu & Bentler, 1999). Nested models are evaluated by changes in the fit indices: changes in RMSEA less than 0.015 and CFI less than 0.01 indicate good relative fit for metric invariance and scalar invariance while changes in SRMR less than 0.03 and 0.01 indicate good relative fit for metric invariance and scalar invariance respectively (Chen, 2007). Although these benchmarks are based on maximum likelihood estimation, they are used as guides in the analysis. The results are also checked using Jrule for Mplus 0.91 (Oberski, 2008). Jrule (judgement rule) for Mplus is a programme that takes the Mplus output as its input and it uses a combination of the expected parameter change, modification index and power (all obtained or calculated automatically from the Mplus output) to detect parameter misspecifications which can occur in spite of adequate global fit as indicated by the fit indices (Saris, Satorra, & van der Veld, 2009; Van der Veld & Saris, 2011b). In this case, high power is set at 0.80 and Type I error at 0.05. The misspecification is set to 0.10 for error covariances and at 0.40 for factor loadings.

5.5 Results

5.5.1 RSs

The Cronbach alpha reliability for ARS, ERS DARS and MRS are 0.88, 0.83, 0.68 and 0.95 respectively in the urban group and 0.74, 0.78, 0.58 and 0.88 respectively in the rural group. These values are calculated from the three indicators computed for each RS and they are relatively high except the alpha for DARS. The corresponding standardised factor loadings of the indicators are all larger than 0.70 except for one loading of ARS in the rural group (0.65) and the loadings on DARS which range between 0.50 and 0.68. The results for DARS should therefore be interpreted with caution.

Although the four-factor RIRSMACS model fits the data well in the rural (RMSEA=0.03, CFI=1.00, SRMR=0.04) and the urban (RMSEA= 0.06, CFI=0.99, SRMR=0.04) group, ARS lacks discriminant validity since the square root of the average variance extracted is lower than the correlation of ARS with ERS (Fornell & Larcker, 1981). The correlations between ARS and ERS are 0.74 and 0.88 whereas the square roots of the average variance extracted for ARS are 0.70 and 0.83 in the rural and urban group respectively. As a result, we drop ARS from the analysis. This decision is supported by two additional observations. Firstly, we observe in further analysis that the rural–urban, mean difference in ARS is significant only at the 10% level. However, the RSs are expected to affect comparability if the levels of the RSs factors differ between the groups (Weijters et al., 2008). Secondly, including ARS in the model results in a non-positive definite matrix in one part of the analysis. This would lead to biased standard errors (McQuitty, 1997).

The revised, three-factor, RIRSMACS model (containing ERS, DARS and MRS) fits the data adequately in both groups (see Table 5.3). Nevertheless, the factor convergent validity, measured by the average variance extracted, for DARS is disappointing: 0.31 (rural); and 0.41 (urban). These values are

Table 5.3: Fit Statistics and Indices for the RSs Models

	χ^2	df	$\Delta\chi^2$	Δdf	$RMSEA$	$\Delta RMSEA$	CFI	ΔCFI	$SRMR$	$\Delta SRMR$
Rural	28.76*	15			0.02		1.00		0.02	
Urban	66.55*	15			0.08		0.98		0.04	
CF	95.31*	30			0.05		0.99		0.03	
MET	135.86*	36	40.55*	6	0.06	0.005	0.98	−0.007	0.04	0.006
SC	163.35*	42	27.49*	6	0.06	0.002	0.98	−0.004	0.04	0.002

* indicates significance at the level. + significant change in fit index. CF – configural invariance. MET – metric invariance. SC – scalar invariance.

Table 5.4: Standardised Factor Loadings for Political Cynicism

Item Code	RSs not Controlled		RSs Controlled	
	Rural	Urban	Rural	Urban
CYN1	0.54	0.65	0.54	0.66
CYN2	0.75	0.89	0.73	0.87
CYN3	0.67	0.67	0.65	0.64
CYN4	0.69	0.55	0.67	0.53
Average variance extracted	0.44	0.49	0.42	0.47

both below the recommended 0.50 benchmark (Fornell & Larcker, 1981). This is consistent with the low alpha reliability levels and it reinforces the need for caution in the interpretation of the results for DARS. On the other hand, discriminant validity is achieved for each factor since the square root average of the variance extracted for each factor exceeds the correlations among the factors (Fornell & Larcker, 1981).

The RSs models exhibit both metric and scalar invariance (also confirmed by Jrule) and their means may therefore be compared between the rural and urban groups. The results indicate that ERS (mean=0.07, SE=0.02, t=3.64, effect size =0.37), DARS (mean=0.08, SE=0.01, t=7.33, effect size=0.76) and MRS (mean=0.03, SE=0.01, t=2.34, effect size=0.22) are higher in the urban group. The sizes of these effects (mean difference divided by standard deviation in the rural group) are not very large, but the RSs are expected to differentially bias the measurements between the two groups. We examine this by comparing the measurements of the attitude constructs.

5.5.2 Political Cynicism

Although the RMSEA is somewhat large, the model without RSs controlled fits adequately in both groups (Table 5.5). The factor has low convergent validity, but the loadings are still relatively large (Table 5.4). With these models accepted as fitting adequately without modification, configural invariance is achieved.

The test for metric invariance fails as indicated by large changes in the CFI and SRMR. The difference in Chi-square is also significant (Table 5.5). Assisted by the modification indices and expected parameter change, we determine that the loading of the second item (Parties are only interested in my vote, not in my opinion) is lower in the rural group. Rural and urban residents therefore appear to interpret this item differently. When this loading is estimated freely, an adequately fitting partial metric invariance model is

Table 5.5: Fit Statistics and Indices for Political Cynicism

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	$RMSEA$	$\Delta RMSEA$	CFI	ΔCFI	$SRMR$	$\Delta SRMR$
RSs not Controlled										
Rural	19.12*	2			0.08		0.96		0.03	
Urban	13.98*	2			0.08		0.98		0.03	
CF	33.10*	4			0.08		0.97		0.03	
MET	52.51*	7	19.41*	3	0.07	0.004	0.96	0.015^{+}	0.06	-0.033^{+}
PMET	41.49*	6	8.39*	2	0.07		0.97		0.05	
PSC1	91.47*	8	49.98*	2	0.10	0.031^{+}	0.91	-0.056^{+}	0.07	0.020
PSC2	42.07*	7	0.59*	1	0.06		0.97		0.05	
RSs Controlled										
Rural	130.73*	50			0.04		0.96		0.04	
Urban	180.29*	50			0.06		0.98		0.06	
CF	378.76*	112			0.05		0.96		0.05	
MET	398.06*	115	19.30*	3	0.05	0.001	0.96	−0.002	0.06	0.004
SC	444.64*	118	46.58*	3	0.06	0.004	0.95	−0.008	0.06	0.004

* indicates significance at the level. + significant change in fit index. CF–Configural invariance. MET – metric invariance. PMET – partial metric invariance. SC – scalar invariance. PSC –partial scalar invariance.

obtained.

Since only partial metric invariance is achieved, only partial scalar invariance may be evaluated. The first partial scalar invariance model fits poorly (see Table 5.5) due to a larger intercept of the third item (Most politicians promise a lot, but don't do anything) in the urban group. With this intercept freed, a close fit to the partial metric invariance model is achieved. Given these modifications, only the first and fourth items contribute to the comparison of the factor means. The results indicate that urban residents are less cynical about politics and the effect size, obtained by dividing the mean difference by the standard deviation in the rural group, is moderate (Table 5.6).

Controlling the impact of the RSs results in small changes to the factor loadings and reductions in the average variances extracted of approximately 5% and 4% in the rural and urban groups respectively (Table 5.4). Notably, full MI is achieved without any modifications (see Table 5.5) and all the items now contribute to the difference between the factor means. The mean difference remains significant and negative sign confirms that urban residents are less cynical about politics (Table 5.6). The corrections for the RSs also result in a drop in the effect size of approximately 23%. In spite of the small impacts on the factor loadings, the RSs hinder MI and bias the comparison of the factor means.

Evaluating the models by focusing on detecting misspecifications rather than on the fit indices using Jrule for Mplus 0.91 (Saris et al., 2009; Van der Veld & Saris, 2011b) leads to specification of two error covariances between the indicators of cynicism: between the third and fourth indicators in the rural group (modification index=7.47, expected change=0.16, power=0.39), and between the second and third indicators in the urban group (modification index=7.19, expected change=0.24, power=0.20). With these changes admitted, full metric invariance is achieved, but only partial scalar invariance due to non-invariance of the third indicator is again observed when the RSs are

Table 5.6: Mean Rural–Urban Differences in Cynicism

Method of Evaluation	Factor	RSs Controlled	Mean Difference	SE	t	Effect Size
Fit Indices	Cynicism	No	−0.45*	0.08	−5.87	0.66
	Cynicism	Yes	−0.35*	0.09	−3.77	0.51
Identifying Misspecifications	Cynicism	No	−0.40*	0.08	−5.17	0.57
	Cynicism	Yes	−0.44*	0.10	−4.34	0.60

* significant at the 5% level. The rural group is the baseline for comparison. The effect size is obtained by dividing the mean difference by the standard deviation in the rural group.

not controlled. With the RSs controlled, both full metric and full scalar invariance are again achieved and the mean difference in cynicism also remains significant and negative in both models (Table 5.6). However, the effect size of this difference is larger when the models are evaluated with this approach even though the effect of controlling the RSs is less pronounced.

The two sets of results demonstrate that the method of model evaluation can affect the results obtained (Saris et al., 2009; Van der Veld & Saris, 2011b). In spite of the this, we note that the results for MI are still similar with respect to invariance of the item intercepts, i.e., the RSs hinder scalar invariance.

5.5.3 Perceived Discrimination, Economic Insecurity and Social (Dis) Trust

Perceived discrimination, economic insecurity and social (dis)trust are evaluated simultaneously. In the model, the two item loadings on (dis)trust are set to 1 due to a negative residual variance of the second indicator in the urban group.

The models fit the data adequately both without and with the RSs controlled (Table 5.8). Although, the average variance extracted for each factor in each group is adequate (at least 0.50) both without and with the RSs, the size of the factor loadings and the convergent validity of the factors change substantially when the RSs are controlled (Table 5.7). The average variance extracted for discrimination, insecurity and distrust decrease by approximately 34%, 18% and 33% respectively in the rural group and by 12%, 6% and 20% respectively in the urban group. RSs therefore inflate the factor loadings of most of the items and the impact is more pronounced in the rural group. The effect of the RSs is most severe for perceived discrimination and least severe for economic insecurity. It is important to note here that although the factor loadings are affected substantially, full metric and full scalar invariance are achieved both without and with the RSs controlled (Table 5.8).

Table 5.7: Standardised Factor Loadings for Perceived Discrimination, Economic Insecurity and Social (Dis)Trust

Item Code	RSs not Controlled						RSs Controlled					
	Rural			Urban			Rural			Urban		
	DISC	INSE	DIST	DISC	INSE	DIST	DISC	INSE	DIST	DISC	INSE	DIST
DISC1	0.80			0.80			0.70			0.73		
DISC2	0.94			0.94			0.85			0.89		
DISC3	0.87			0.87			0.54			0.82		
INSE1		0.91			0.91			0.81			0.88	
INSE2		0.88			0.88			0.79			0.86	
INSE3		0.68			0.68			0.64			0.65	
DIST1			0.83			0.83			0.71			0.74
DIST2			0.85			0.85			0.66			0.76
AVE	0.76	0.69	0.71	0.76	0.69	0.71	0.50	0.56	0.47	0.67	0.65	0.56

DISC – Perceived Discrimination. INSE – Economic Insecurity. DIST – Social (Dis)Trust. AVE – Average variance extracted.

Table 5.8: Fit Statistics and Indices for Perceived Discrimination, Economic Insecurity and Social (Dis)Trust

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*RMSEA*	$\Delta RMSEA$	*CFI*	ΔCFI	*SRMR*	$\Delta SRMR$
RSs not Controlled										
Rural	54.19*	18			0.03		0.98		0.04	
Urban	48.04*	18			0.04		0.98		0.04	
CF	102.24*	36			0.04		0.98		0.04	
MET	119.13*	40	16.89*	4	0.04	0.001	0.98	−0.002	0.05	0.011
SC	129.73*	45	10.60	5	0.04	0.000	0.98	−0.002	0.05	0.002
RSs Controlled										
Rural	196.53*	96			0.02		0.98		0.04	
Urban	309.91*	96			0.06		0.95		0.05	
CF	573.97*	204			0.04		0.96		0.04	
MET	593.58*	208	19.61*	4	0.04	0.001	0.96	−0.002	0.05	0.003
SC	603.81*	213	10.23	5	0.04	−0.001	0.96	0.000	0.05	0.000

* Significant at the 5% level. All changes in fit indices lack significance. CF – Configural invariance. MET – metric invariance. SC – scalar invariance.

Table 5.9: Mean Rural–Urban Differences in Perceived Discrimination, Economic Security and Social (Dis)Trust

Method of Evaluation	Factor	RSs Controlled	Mean Difference	SE	t	Effect Size
Fit Indices	Perceived Discrimination	No	−0.09	0.06	−1.34	0.19
	Economic Insecurity	No	0.13	0.09	1.47	0.13
	Social (Dis)Trust	No	0.00	0.06	−0.05	0.01
	Perceived Discrimination	Yes	−0.29*	0.08	−3.60	0.46
	Economic Insecurity	Yes	0.02	0.11	0.12	0.02
	Social (Dis)Trust	Yes	0.08	0.06	1.32	0.16
Identifying Misspecifications	Perceived Discrimination	Yes	−0.28*	0.08	−3.57	0.41
	Economic Insecurity	Yes	0.02	0.11	0.15	0.02
	Social (Dis)Trust	Yes	0.08	0.06	1.31	0.16

* significant at the 5% level. The rural group is the baseline for comparison. The effect size is obtained by dividing the mean difference by the standard deviation in the rural group.

Before the RSs are controlled, there are no significant mean rural–urban differences in the factors (Table 5.9). When the RSs are controlled, perceived discrimination is significantly lower in the urban group whereas economic insecurity and social (dis)trust continue to show no significant rural–urban differences (Table 5.9).

When evaluated using Jrule, full metric and scalar invariance are achieved without controlling the RSs. When the RSs are controlled, the equality constraint on the third indicator of perceived discrimination (DISC3) appears to be misspecified (modification index=13.29, expected change=-0.18, power=0.22, rural group). This leads to partial metric invariance and subsequently to partial scalar invariance. However, consistent with the results based on evaluation with the global fit indices, a previously absent mean difference in perceived discrimination emerges when the RSs are controlled and the direction of the difference is also the same (Table 5.9).

Two sets of conclusions may be drawn from this set of results. Based on the global fit indices, it is reasonable to conclude that RSs mask a moderate mean difference in perceived discrimination even though full MI is demonstrated. On the contrary, the approach of detecting misspecifications, indicates that the RSs result in full metric invariance by differentially inflating the loading of the third perceived discrimination item (more in the rural group) and that they also masked a moderate mean difference in perceived discrimination.

5.6 Discussion

This study confirms that ERS, DARS and MRS are higher in urban areas and that these differences result in biased measurements and substantial differences in research results. RSs inflate factor loadings differentially between rural and urban regions and hence affect factor convergent validity. They also differentially affect item intercepts. As a consequence, RSs can either distort the effect sizes of factor mean differences or conceal mean differences altogether.

Therefore, RSs bias measurements and research results across rural and urban areas just as they do across data collection modes (Weijters et al., 2008) and across countries (Kankaraš & Moors, 2011; Welkenhuysen-Gybels et al., 2003). Hence, it is as important to control RSs when data are collected across rural and urban areas as in multimode or cross-cultural setting.

While the effect sizes of the RSs differentials between rural and urban regions are not very large, they are at least as large as those between modes of data collection. Consequently, the ecological RSs bias is at least as severe in data that are pooled across rural and urban areas as in multimode data (Thomas et al., 2014). Pooling data assumes preservation of the meaning of the items across the groups (F. J. R. Van de Vijver & Poortinga, 1997; F. J. R. Van de Vijver & Tanzer, 2004), but this assumption is untenable when data are distributed across rural and urban areas since RSs bias may be present.

The effect of rural–urban RSs bias is not the same for all constructs. This is not surprising given that RSs are expected to exhibit construct specificity (Billiet & McClendon, 2000). On the one hand, we find that while the factor loadings for political cynicism do not change substantially, at least scalar invariance is affected and the effect size of the factor mean difference changes. On the other hand, the factor loadings for perceived discrimination, economic insecurity and social (dis)trust change substantially and differentially between the groups when the RSs are controlled. Some constructs may therefore be more substantially affected than others and it is difficult to determine the extent of the bias beforehand. It is therefore important to ensure that RSs bias does not enter any part of the analysis.

An important consideration when measuring and controlling for RSs is the confounding of content with style which occurs if the RSs items are the same as those measuring the substantive content or if the items measure some common factor (Mõttus et al., 2012; Van Vaerenbergh & Thomas, 2013). This can be avoided by using a random selection of items which measure different underlying constructs and which are uncorrelated (Weijters et al., 2008). This study uses the RIRSMACS model with 27 items which measure different constructs and which have a low average interitem correlation. However, corrections for RSs with this method can be done with as few as 6 items although 15 is recommended (Weijters et al., 2008). There are also other methods of correcting for RSs, which do not require additional items. For example, the style factor approach only requires reversed items to control ARS (Billiet & McClendon, 2000). In addition, standardisation which has none of these requirements is also used to control RSs (R. Fischer, 2004). There are several options from which researchers may choose, but it is important to understand their advantages and disadvantages (see Van Vaerenbergh & Thomas, 2013, for a review of the methods).

Gans (1962) suggests that cultural differences between rural and urban areas are more likely in more traditional societies. This study targets a developing country and confirms the adverse effects of rural–urban RSs differentials on measurements and research results. However, given that there is evidence of rural–urban subcultures in more developed societies (Tittle & Grasmick, 2001), the possible differential ecological effects of RSs on substantive research outcomes in Western societies should not be ignored.

In general, MI is required for group comparisons, but it does not guarantee the absence of RSs bias between countries (Welkenhuysen-Gybels et al., 2003). This paper extends the relevance of this conclusion to within-country, rural–urban comparisons. In the analysis, no rural–urban, mean differences in perceived discrimination, economic insecurity and social (dis)trust are detected before controlling the RSs, but perceived discrimination is higher in the rural group when the RSs are controlled. MI therefore provides no guarantee against RSs bias between rural and urban regions. RSs can also hinder MI (Kankaraš & Moors, 2011) as observed for cynicism or they may result in MI as observed for perceived discrimination when model evaluation is based on Jrule. Ignoring RSs has non-trivial consequences for MI itself even in within-country research. In particular, the greater difficulty in achieving higher levels of MI (Horn & McCardle, 1992; Schmitt & Kuljanin, 2008) is at least partially due to RSs, but RSs may also be the cause of higher levels of MI.

In light of the consequence of rural–urban RSs differentials for measurement comparability and the ineffectiveness of traditional MI evaluations in detecting this bias, controlling RSs while demonstrating MI should become a basic research requirement. Controlling RSs is necessary in both within-country and cross-cultural research. In addition, existing theories should be re-examined with RSs controlled (Moors, 2012).

5.7 Limitations and Recommendations for Future Research

This study employs the RIRSMACS model which necessitates Confirmatory Factor Analysis (CFA) which may produce different results for MI compared to other modelling techniques (Kankaraš, Vermunt, & Moors, 2011) as may the RIRSMACS model compared to other approaches for RSs (Van Vaerenbergh & Thomas, 2013). The effects of the rural–urban RSs divide should therefore be investigated with alternative methodologies such as the Style Factor with CFA (Billiet & McClendon, 2000), Latent Class Confirmatory Factor Analysis (Kieruj & Moors, 2010, 2013; Moors, 2004, 2012) and Item Response Theory (see Jin & Wang, 2014). This will help in determining whether the effects are consistent across methods. These studies should encompass several content areas in both Western and non-Western countries to provide enough evidence to convince researchers to control RSs in within-country research and to assist in determining the generalizability of the rural–urban, RSs divide to Western contexts. Finally, researchers should investigate the effects of the rural–urban RSs divide on structural parameters.

References

Abts, K. (2012). *Maatschappelijk onbehagen en etnopopulisme: Laatmoderniteit, burgers, ressentiment, migranten, politiek en extreemrechts* (Unpublished doctoral dissertation). Faculty of Social Sciences, University of Leuven.

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Deerfield, IL: AAPOR.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, *66*(3), 374–392.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement in variance. *Psychological Bulletin*(3), 456–466.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155–175.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? the impact of making inappropriate comparisons in cross-cultural research.

Journal of personality and social psychology, *95*(5), 1005–1018.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*(1), 1–27.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149.

Fischer, C. S. (1975). Toward a subcultural theory of urbanism. *American Journal of Sociology*, *80*(6), 1319–1341.

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, *35*(3), 263–282.

Fornell, C., & Larcker, D. F. (1981). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Journal of Marketing Research*, *18*, 39–50.

Gans, H. (1962). Urbanism and suburbanism as ways of life: A re-evaluation of definitions. In A. M. Rose (Ed.), *Human behavior and social processes; an interactionist approach* (pp. 625–648). Boston: Houghton-Miffl.

Horn, J. L., & McCardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Jin, K.-Y., & Wang, W.-C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement*, *74*(1), 116–

138.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*(2), 264–277.

Kankaraš, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija*, *43*(2), 121–136.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, *40*(2), 279–310.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*(3), 320–342.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Locke, K. D., & Baik, K.-D. (2009). Does an acquiescent response style explain why Koreans are less consistent than Americans? *Journal of Cross-Cultural Psychology*, *40*(2), 319–323.

McQuitty, S. (1997, January). Effects of employing ridge regression in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(3), 244–252.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, *57*(2), 289–311.

Meuleman, B., Davidov, E., & Billiet, J. (2009). Changing attitudes toward immigration in Europe, 20022007: A dynamic group conflict theory approach. *Social Science Research*, *38*(2), 352–365.

Millsap, R. E. (1995). Measurement invariance predictive invariance and the duality paradox. *Multivariate Behavioral Research*, *30*, 577–605.

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single factor case. *Psychological methods*, *2*, 248–260.

Millsap, R. E. (2007). Invarince in measurement and prediction revisited. *Psychometrika*, *72*, 461–473.

Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5–9.

Millsap, R. E., & Yun-tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479–515.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, *20*(4), 303–320.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Mõttus, R., Allik, J., Realo, A., Rossier, J., Zecca, G., Ah-Kion, J., … Johnson, W. (2012). The effect of response style on self-reported conscientiousness across 20 countries. *Personality & social psychology bulletin*, *38*(11), 1423–36.

Oberski, D. (2008). Jrule for Mplus. Retrieved from `https://github.com/daob/JruleMplus/wiki`

Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Formal definitions of measuement bias and explanation bias clarify measurement and conceptual perspectives. *Journal of Clinical Epidemiology*, *62*, 1126–1137.

Petković, J. (2007). Traditional values and modernization challenges in forming urban and rural culture. *Philosophy, Sociology and Psychology*, *6*, 23–39.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582.

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *94*, 347–363.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*(1), 50–61.

Smith, T. W. (2011). Refining the total error perspective. *International Journal of Public Opinion Research*, *23*, 464–484.

Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, *43*(4), 599–616.

Swyngedouw, M., Abts, K., & Rink, N. (2009). *Belgian general election study 2007. Codebook: Questions and frequency tables.* Leuven: ISPO-K.U.Leuven.

Thomas, T. D., Abts, K., & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement*, *74*(1), 97–115.

Tittle, C. R. (1989). Influences on urbanism: A test of predictions from three perspectives. *Social Problems*, *36*(3), 270–288.

Tittle, C. R., & Grasmick, H. G. (2001). Urbanity: Influences of urbanness, structure, and culture. *Social Science Research*, *30*(2), 313–335.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Van der Veld, W. M. (2008). Judging misspecifications in structural equation models using JRule. *Symposium on Testing Structural Equation Models at the ISA-RC33, 7th International Conference on Social Science Methodology. Naples (Italy), September 1–5*, 1–5.

Van der Veld, W. M., & Saris, W. E. (2011a). Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Methods and applications in cross-cultural analysis* (pp. 117–148). New York: Taylor & Francis Group.

Van der Veld, W. M., & Saris, W. E. (2011b). Cause of generalized social trust. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 207–247). New York: Routledge: Taylor & Francis Group.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana*. Ghent, Belgium: Ghent University.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications Inc.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*, 29–37.

Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, *54*(2), 119–135.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, *44*(1), 1–24.

CHAPTER 6

MEASURING INSTITUTIONAL TRUST IN GUYANA: A SECOND-ORDER FACTOR MODEL WITH CORRECTIONS FOR RESPONSE STYLES

Measuring Institutional Trust in Guyana: A Second-Order Factor Model with Corrections for Response Styles

Troy Devon Thomas[1,2], Koen Abts[3], Koenraad Stroeken[1] and Patrick Vander Weyden[4]

[1] *Department of Comparative Science of Cultures, Ghent University.*

[2] *Department of Mathematics, Physics and Statistics, University of Guyana.*

[3] *Centre for Sociological Research, University of Leuven*

[4] *Founder of the independent research company — FocusUP.*

Abstract

Institutional trust is often measured by several separate items and by sum scores. Whereas the results of individual-items analyses are difficult to summarise, sum scores may be meaningless due to untenable assumptions about the dimensions of the construct. The use of sum scores in less advanced democracies is often based on the assumption that institutional trust is unidimensional; however, factor analysis with data from Guyana indicates that this assumption is untenable. Based on Guyanese data, we propose a second-order factor model for institutional trust. However, even with a factor model, we find that research results are still affected by response styles (RSs). RSs inflate item validity and factor convergent validity and may either distort regression coefficients or altogether result in spurious effects in institutional trust research. In order to reduce bias in institutional trust research, factor models with RSs controlled should be used instead of individual-items and sum score analyses.

Keywords: trust, democracy, institutions, response styles, Guyana

---

6.1 Introduction

Institutional trust means that the citizens have confidence in the competence and good intentions of the institutions in situations in which they are vulnerable (Secor & Loughlin, 2005). When trust is high, the authorities are not expected to abuse their powers or intentionally cause harm and the citizens tend to voluntarily defer to decisions and to comply with rules (Khodyakov, 2007). For example, citizens are more likely to pay taxes when institutional trust is high (Hug & Spörri, 2011). The general belief that institutional trust is on the decline (Shlapentokh, 2006) has motivated much research on the topic. However, there is an overwhelming focus on more consolidated democracies which are thought to have a more differentiated view of institutions than less developed societies. In particular, institutional trust is thought to be multidimensional in more advanced democracies, but unidimensional elsewhere (Mishler & Rose, 2001). Consequently, measurement models for institutional trust that are developed in more advanced democracies are unlikely to be inappropriate for less advanced democracies. In this regard, the limited evidence from developing societies is an important shortcoming. The general assumption that institutional trust is unidimensional in less advanced democracies needs to be evaluated further.

A second issue is the measurement of institutional trust. Research results are often based on individual items about particular institutions (for example, Blanco, 2013; Blanco & Ruiz, 2013) or on sum scores over such items (Poznyak, Meuleman, Abts, & Bishop, 2013). Analysing individual items makes it difficult to provide overviews of institutional trust whereas sum scores lead to untrustworthy results (Neale, Lubke, Aggen, & Dolan, 2005; Poznyak et al., 2013). These limitations may be addressed with the use of factor models, but even when factor models are employed, response styles (RSs) which are the respondents' systematic tendencies to respond to rating scale items in certain ways regardless of their content (Van Vaerenbergh & Thomas, 2013) are not controlled. RSs bias research results and can lead to spurious regression relationships (Moors, 2012). Research results for institutional trust and its explanatory variables are therefore subject to bias from a combination of sources.

This paper evaluates a measurement model — factor model — for institutional trust with corrections for RSs using data from Guyana which is an English-speaking developing country in South America. It compares the effects of the respondents' socio-demographic characteristics on institutional trust across four methods of measurement: single items, sum scores, factor model and factor model with corrections for RSs. This contributes to the literature in four main ways. First, by focusing on Guyana which is a less advanced democracy, it adds to what is known about institutional trust in fledgling democracies. Second, the comparisons of the regression effects across the four methods demonstrates the impact of the method of measurement and the impact of RSs on the results of institutional trust research. Third, the procedure employed to correct or RSs in the factor model can serve as a guide to researchers. Fourth, the factor model enables evaluations of the dimensions of institutional trust and facilitates assessment of the validity of the sum score approach.

6.2 Measurement of Institutional Trust

In general, measuring institutional trust with several items focusing on various institutions is recommended because this approach captures the variations across the institutions (Mishler & Rose, 1997). Although this approach is adopted most often, there are three popular ways of analysis such data in the institutional trust literate. The items are often analysed separately, as sum scores or with factor models.

Individual Items. Studying institutional trust with individual items require several separate segments of analyses (see Christensen & Læ greid, 2005). However, the specific institutions are seldom of interest. Combining the results of the separate analyses for separate institutions into a meaningful overview is often difficult. This difficulty is exacerbated when many institutions are included and when regression effects of the same predictor vary between institutions (for example, see Blanco, 2013; Blanco & Ruiz, 2013). In addition, measurement errors are not controlled when individual items are used and this can affect the validity of the results.

Sum Scores. Sum scores over several items referring to various institutions can easily provide an overall summary or summaries by category of institutions (see Chang & Chu, 2006; Hamm et al., 2011; Huang, Lee, & Lin, 2013; Hutchison & Johnson, 2011; Lühiste, 2006). However, a major problem with sum scores is that their use is based on the untested assumption that the items summed measure the same dimension. Institutional trust items are often combined into a single sum score in research in developing democracies thus reflecting the belief that the construct is unidimensional (for example, Mishler & Rose, 2001; Rohrschneider & Schmitt-Beck, 2002). This approach neglects the possible effect of culture on measurements. Culture can affect the structure of measurement models even when the same items are used (Van de Vijver & Poortinga, 1997). Specifically for institutional trust, the functioning of the items are likely to be affected by the purpose for which institutions are set up and the way in which they operate within the specific context. As such, the measurements may differ from one country to another or over time as the conditions within a country change (Bouckaert & Van De Walle, 2001; Poznyak et al., 2013). The assumption that the items form a single dimension may therefore be incorrect and as a consequence, research results that are based on a single sum score for institutional trust may be meaningless.

Apart from the specific issues related to institutional trust research, there are more general problems with sum scores. Sum scores neglect measurement error and regard the entire responses of the individuals as meaningful (Neale et al., 2005). The inherent assumption that the items are perfectly reliable is unlikely to be correct and as such regression estimates become inconsistent (Bollen & Lennox, 1991). When a single predictor is used in regression analysis, the coefficient is likely to underestimate the true value, but when several explanatory variables are included, the direction of the bias cannot be predicted (Bollen & Lennox, 1991). In addition, Shevlin, Miles, and Bunting (1997) indicate that the bias due to sum scores is likely to be downward and more pronounced when the reliability of the items are moderate to low. These effects of sum scores remain even when the dimension(s) that the items are assumed to measure are correct (Neale et al., 2005).

Factor Analysis. Factor models for institutional trust facilitate determination of the dimensions measured by the items while at the same time, measurement error is taken into account. Furthermore, factor models perform well even when the reliability of the items are moderate to low (Shevlin et al., 1997). Factor analysis therefore overcomes the outlined limitations of individual items and sum scores.

The results of exploratory factor analysis show that a single dimension of institutional trust is justifiable in some cases (Listhaug, 1984; Mishler & Rose, 1997, 2005), but that up to three dimensions are appropriate in other cases (Bean, 2003; Rothstein & Stolle, 2008). In particular, Rothstein and Stolle (2008) and Bean (2003) identify the three dimensions of institutional trust as partisan (example, parliament, government), non-partisan (example, police, army) and media. Memberships in partisan institutions is based on elections whereas membership in the non-partisan institutions is not. This potential for different categorisations of the items evidences their differential functioning in different societies. However, Mishler and Rose (1997) explains that a single dimension is appropriate for less consolidated democracies since their citizens lack the democratic sophistication required to discern separate dimensions. In spite of this, we argue for re-evaluations of the measurement models in the particular country under consideration regardless of its stage of democratic development since this provides assurance about the form of the measurement model and avoids the biases associated with the use of individual items and sum scores.

6.3 The Impact of RSs on Measurements

In general, institutional trust is measured with the use of rating scales, but RSs are not controlled in the analysis. As such, research results for institutional trust is subject to RSs bias regardless of the method of measurement employed. The RSs that are studied most often are acquiescence RS (ARS: tendency to agree), extreme RS (ERS: tendency to use the scale endpoints), disacquiescence RS (DARS: tendency to disagree) and midpoint RS (tendency to use the scale midpoint). RSs bias factor loadings and constructs means (Billiet & McClendon, 2000; Kankaraš & Moors, 2011; Weijters, Schillewaert, & Geuens, 2008; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). For example, higher(lower) ERS increases(decreases) factor loadings whereas higher(lower) ARS increases(decreases) the means of manifest variables (Cheung & Rensvold, 2000). These effects are non-uniform across subgroups of respondents. For example, the mean levels of the RSs differ significantly between rural and urban areas in Guyana (Thomas, Abts, & Vander Weyden, 2014) and these rural-urban RSs differentials bias within-country measurement comparability, differentially affect factor convergent validity and can either distort or altogether conceal mean differences between factor means (Thomas, Abts, & Vander Weyden, in press).

RSs also bias structural relationships. Moors (2012) shows that the well-accepted gender effect on leadership styles is really due to RSs. RSs can cause spurious relationships by inflating(deflating) factor variances and covariances in addition to the other effects on the factorial structure of measurement models (Baumgartner & Steenkamp, 2001; Moors, 2012). They also distort the discriminant validity of factors entered into the same model. Research on institutional trust often focuses on regression relationships and may include correlated factors. Since the measurements are not usually adjusted for RSs, the results are likely to be biased even when factor models are employed. This limits confidence in the established relationships between institutional trust and other variables. We therefore argue for the use of factor models with corrections for RSs in institutional trust research.

6.4 Socio-Demographic Determinants of Trust in Institutions

In this paper, we focus on the effects of age, gender, eduction and ethnicity. We investigate the effects of these variables in order to illustrate the impacts of the methods of measurements and the RSs on the results of substantive institutional trust research.

Trust in institutions is associated with the socio-demographic characteristics of the individuals. However, when individual items are used, the associations seem to depend on specific institutions. Furthermore, overall, the associations may depend on the country (Blanco, 2013; Christensen & Læ greid, 2005; Huang et al., 2013). Whereas some find that institutional trust increases with age (Hutchison & Johnson, 2011; Listhaug, 1984), others find no effect of age (Lühiste, 2006; Mishler & Rose, 1997; Rohrschneider & Schmitt-Beck, 2002). Research results for gender are also inconsistent. Listhaug (1984) and Mishler and Rose (1997) indicate that males are less trusting of institutions, whereas Hutchison and Johnson (2011) report that gender has no effect on institutional trust. For education, some find a negative effect on institutional trust (Blanco, 2013; Hutchison & Johnson, 2011; Lühiste, 2006; Rohrschneider & Schmitt-Beck, 2002), but Abts (2012) indicate that this effect is positive and Mishler and Rose (1997) indicate that there is no such relationship. Finally, Lühiste (2006) indicates that the majority ethnic group has higher institutional trust whereas Hutchison and Johnson (2011) find no consistent effect of ethnicity. In spite of the inconsistencies in the results, the respondents' socio-demographic characteristics appear to predict institutional trust in most cases. These variables are therefore expected to provide a basis for evaluating the impact of methods of measurement and the effects of the RSs on structural relationships in institutional trust research.

6.5 Data and Methods

6.5.1 Data

The data used in this study are obtained from the Values and Poverty Study in Guyana (VAPO Guyana) which was conducted between April and May 2012. The study was funded by the Flemish Inter-University Council (VLIR) and jointly executed by the University of Guyana and Ghent University. It investigates both methodological and substantive issues and provides an opportunity to correct for RSs using representative indicators. The data were collected via face-to-face interviews by a survey organisation (DPMC) under the supervision of the University of Guyana and Ghent University. These data are representative of the coastal regions (region 2, 3,4, 5, 6 and 10) which account for approximately 90% of the country's population.

The VAPO Guyana employed a sampling procedure which randomly selected municipalities with probability proportional to size, and respondents within the municipalities with equal probabilities. This resulted in the selection of 87 clusters within 51 municipalities and a total of 1048 individuals were interviewed at a response rate of 87% (American Association for Public Opinion Research [AAPOR] RR2; AAPOR, 2011). The data are weighted for nonresponse using iterative proportional fitting.

Socio-Demographic Characteristics. The socio-demographic characteristics included in this study are age, gender, education and ethnicity. Age is measured in years and it has an average of 36.25 years. Males account for approximately 49% of the sample and the category female is used as the reference group. Education has three levels: primary or lower (Low Education), secondary (reference group) and higher than secondary (High Education). Approximately 30.5%, 60% and 12.5% of the sample has up to primary, secondary and higher than secondary education respectively. Ethnicity is coded dichotomously to reflect the majority (East Indians: 46%) versus the remainder of the population (Afros, Amerindians, Chinese, Portuguese and White). The combined minority ethnicities is used as the reference group.

6.5.2 Methods

A combination of Ordinary Least Squares Regression (OLSR), confirmatory factor analysis (CFA) and structural equation modelling are used to analyse the data. The OLSR models are estimated with IBM SPSS Statistics 21 and the CFA and structural equation models (SEM) are implemented with Mplus 7.11 with robust maximum likelihood estimation. Given the large sample size, the CFA models are evaluated with alternative fit indices (Chen, 2007).[1] However, Jrule is also used to identify misspecification that may go undetected by the global fit indices (Oberski, 2008; Van der Veld, 2008).[2] In addition, the convergent validity and discriminant validity of the factors are evaluated. The convergent validity factors are judged to be adequate if the average variance extracted (AVE) is greater than or equal to 0.50 whereas discriminant validity is adequate if the AVE($\sqrt{AVE}$) for the factor exceeds its

[1] The Root Mean Square Error of Approximation (RMSEA) less than or equal to 0.06, Comparative Fit Index (CFI) and Tucker and Lewis Index (TLI) greater than or equal to 0.95 and Standardised Root Mean Square Residual (SRMR) less than or equal to 0.05 are indicative of acceptable overall fit (Byrne, 2012; Hu & Bentler, 1999).

[2] Jrule (judgement rule) for Mplus is a programme that takes the Mplus output as its input and it uses a combination of the expected parameter change (EPC), modification index and power (all obtained or calculated automatically from the Mplus output) to detect parameter misspecification which can occur in spite of adequate global fit as indicated by the fit indices (Saris, Satorra, & van der Veld, 2009; Van der Veld & Saris, 2011). In this study, high power is set at 0.80 and Type I error at 0.05. The EPC is set to 0.10 for error covariances and at 0.40 for factor loadings.

covariances(correlations) with the other included factors (Fornell & Larcker, 1981).

Measurement of Institutional Trust. Institutional trust is measured by items which capture the respondents' ratings of their levels of trust in nine different institutions in Guyana. The item reads as follows: *"Can you tell me to what extent you trust the following institutions?"* The institutions presented are: *the justice system, Guyana Defence Force (army), parliament, national government, Guyana Police Force (police), national elections, political parties, mayor's office of your city or town/neighbourhood democratic council (NDC) chairman's office, and the Regional Democratic Council (RDC).* Trust in each institution is rated on a five-point scale which is has both numeric and verbal labels: 1 (distrust very much), 2 (distrust), 3 (neither trust nor distrust), 4 (trust), 5 (trust very much).

A step-wise approach is followed in the analysis. First, the individual items are analysed separately. In this case, nine separate OLSR models are estimated; one for trust in each institution. Second, the sum score analysis is done using OLSR. Three separate sum score analyses are conducted. The first of is an overall sum calculated across the nine items under the assumption that all the items form a single dimension. In addition to this, two other sum score are analysed. In these remaining two cases, the items are combined to be consistent with the dimensions identified in the next step of the analysis in which a factor model is implemented. In reality, we could not have known about these two dimensions without first going on to use factor analysis. However, they provide a basis for evaluating the sum score approach under the assumption that the implied dimensions are correct. As such, it is judged to be important to include them. These two final sum scores are presented along with the overall sum score before the factor models are presented since this gives a better organisation to the presentation of the results. Third, the data are analysed using a CFA (and SEM) model for institutional trust. Finally, the measurements are adjusted for RSs and re-analysed using CFA (and SEM). This final step involves the suggested model for institutional trust.

To measure institutional trust using CFA, we begin with a one-factor model. This model is expected to hold in less advanced democracies (Mishler & Rose, 2005). If this initial model fails, an alternative is developed through inspection of the modification indices and expected parameter changes in combination with the indications provided by Jrule. The factor models are then extended into SEMs with the respondents' socio-demographic characteristics as the predictors.

Corrections for RSs. Several model-based approaches for correcting for RSs are available (see Van Vaerenbergh & Thomas, 2013, for a review). The use of CFA for institutional trust narrows the potential models to those that can be implemented with CFA. In this regard, Van Vaerenbergh and Thomas (2013) recommend the representative indicators response styles means and covariance structure (RIRSMACS) model (Weijters et al., 2008) because it includes several RSs simultaneously. The RIRSMACS model includes ARS, ERS, DARS and MRS as latent variables, each having three indicators calculated from three blocks of items (one indicator each per block) (Weijters et al., 2008) and the it is flexible enough to permit inclusion or exclusion of various RSs.

In estimating the RSs, the content of the items must be controlled to avoid confounding with style. This is facilitated by the VAPO Guyana which avails dedicated RSs items. Based on a pre-test with 1000 students at the University of Guyana, 35 items (with low correlations; $r \leq |0.3|$) were selected and included in the questionnaire of the VAPO Guyana as dedicated RSs items. These items represent a randomly selection from various constructs covering several topics (including government, politics, society, crime gender roles and many more). The RSs items were then included in the VAPO Guyana questionnaire along with the items designed to measure several substantive constructs.

This was done to ensure that separate items are always available to measure and correct for RSs (Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012).[3] In this study, the RSs are measured by a random selection of 27 of the 35 items (See Appendix A.2) and they have an average interitem correlation of 0.06. The RSs items are all scored on 5-point fully labelled rating scales with disagree/agree verbal labels. The numeric labels of the RSs items match those of the trust items, but the mismatch of verbal labels is a limitation of this study since the scale format can affect RSs (Weijters, Cabooter, & Schillewaert, 2010).

To obtain the values of the RSs indicators, the pool of 27 items is divided at random into three blocks of 9 items each and one indicator per RS is calculated from each block. These values are calculated as:

$$ARS = [f(4) + 2 * f(5)]/k$$

$$ERS = [f(1) + f(2)]/k$$

$$DARS = [2 * f(1) + f(2)]/k$$

and

$$MRS = f(3)/k$$

where $f(x)$ is the frequency of the response option x and k is the number of items per block (Weijters et al., 2008).

In the RIRSMACS model, the indicators that are calculated from the same block of items are all correlated (Weijters et al., 2008). Each indicator of the substantive constructs is also modelled as an indicator of each RS and the substantive construct(s) is(are) not allowed to correlate with the RSs (see Figure 6.2). The impacts of a single RS on the items measuring a particular

[3]Neither pretesting nor separate items are absolutely necessary for controlling RSs. Usually, heterogeneous item can be obtained by randomly selecting one item per construct from the constructs that are included the questionnaire provided that the questionnaire covers a variety of topic areas.

substantive construct are equated, but the effects of different RSs on the same set of items are allowed to be different (Weijters et al., 2008). Given that each of the content items measure aspects of institutional trust, the impact of each RS is set equal for all the items regardless of whether or not they form different factors, but the impacts of separate RSs are allowed to be different.

6.6 Results

6.6.1 Level of Trust

The level of trust in institutions in Guyana is generally low (see Table 6.1). The responses to the items are coded such that lower values indicate lower trust. As observed, only three of the items have scores that are on average above the scale midpoint whereas the other six institutions have average scores that are below the scale midpoint. Notably, the army is the most trusted institution whereas the police is the least trusted. Both of these are non-partisan institutions in the sense that membership is not based on elections.

Table 6.1: Levels of Trust in Institutions

Institution	Mean	Standard Deviation
Justice System	2.96	1.04
Army	3.34	0.98
Parliament	3.17	0.95
National Government	3.06	1.03
Police	2.66	1.11
National Elections	2.89	1.10
Political Parties	2.93	0.96
Mayor's or NDC Office	2.95	1.01
RDC	2.98	1.02

6.6.2 Socio-Demographic Determinants

Individual-Item Analysis. In the individual-item analysis, the ratings for each institution are regressed on the respondents' socio-demographic characteristics using the method of least squares (Table 6.2). As expected, overall interpretations of the results are difficult to deduce because their effects show institution specificities.

Age. Older individuals have less trust in the national government, but age does not predict trust in any of the other institutions.

Gender. Males have less trust in the justice system, national elections and the Mayor's/NDC office than females, but gender is not associated with trust in the other institutions.

Education. Education is significantly related to trust in the parliament, national government, police, national elections, political parties and the RDC, but not the justice system, army and the Mayor's/NDC office. In particular, compared to secondary education, those with higher education are less trusting whereas those with lower education are more trusting of the national government and the police. For some institutions, either one of the higher education or the lower education group is distinguished from secondary education group but not both. This occurs for trust in parliament and the RDC which are lower for more highly educated individuals and for trust in the national elections and political parties which are higher among the low educated individuals.

Ethnicity. The results for ethnicity are much clearer and more easily generalised compared to the other variables. Ethnicity explains trust in each institution. The sign of the coefficients indicate consistently that institutional trust is higher among the majority ethnic group. At the time of the data collection, the incumbent – People's Progressive Party/ Civic — was the political party that is generally thought to be more strongly linked to the majority ethnic group — East Indians — than the minority groups. The positive association between trust and the politically relevant ethnicity suggests that party ideology plays an important role in trust in the institutions in Guyana.

The explained variances of the single-item models show large variations from one institution to the other. For example, approximately 19% of the variance in trust in the national elections is explained whereas only approxi-

Table 6.2: Standardised Coefficients of the Single-Item Regression Models

Institution	Predictor					R-Squared
	Age	Gender	High Education	Low Education	Ethnicity	
Justice System	−0.03	−0.08***	−0.05	0.02	0.25***	0.08
Army	0.05	0.00	−0.05	−0.05	0.11***	0.02
Parliament	0.00	−0.04	−0.08**	0.02	0.21***	0.06
National Government	−0.06*	−0.03	−0.06**	0.10***	0.37***	0.18
Police	−0.06	−0.04	−0.09***	0.08***	0.27***	0.11
National Elections	−0.04	−0.05*	0.02	0.11***	0.40***	0.19
Political Parties	0.00	−0.02	−0.04	0.08**	0.30***	0.12
Mayor's/ NDC Office	−0.06	−0.07**	−0.03	0.05	0.24***	0.07
RDC	−0.02	−0.05	−0.07**	0.02	0.24***	0.07

*, ** and *** ⇒ significance at 10%, 5% and 1% respectively. Gender: ref.– female. Ethnicity: ref.– minority. Education: ref.– secondary.

mately 2% of trust in the army is explained. Noteworthy is the fact that the institution with the highest level of trust (the army) has the lowest explained variance by the socio-demographic characteristics of the respondents. A similar conclusion could be made for the justice system, parliament, Mayors'/NDC office and the RDC for which the explained variances are quite small. However, socio-demographics explain larger proportions of the variances in trust in the national government, national elections and political parties which are all national level partisan institutions.

Sum Score Analysis. A major issue in using the sum score is to determine the number of constructs measured by the items and more specifically which items measure which constructs. Following Mishler and Rose (1997), who argue that different the different items measure the same construct in the less advanced democracies, researchers are likely to use the sum of all the items as a measure of generalised institutional trust. This approach is evaluated here. However, the findings based on factor analysis (presented subsequently) are used to provide two alternative sum score measures. Specifically, the trust items are found to measure two constructs: trust in national institutions and trust in local institutions. The final two items in Table 6.1 measure trust in local institutions whereas the other items measure trust in national institutions. As such, three sets of results are provided for the sum scores.

The internal consistencies of the sum score measures are high: national institutional trust, 0.88; local institutional trust, 0.90; and generalised institutional trust, 0.91. In addition, the socio-demographic variables explain approximately 16%, 8% and 16% of the variance in national, local and generalised institutional trust respectively (Table 6.3).

Age. Age does not predict national, local or generalised institutional trust. In the analysis of the individual items, age predicts only two of the nine trust variables. As such, it is not altogether surprising that it has no effect on these aggregate measures.

Table 6.3: Standardised Regression Coefficients for the Effects of the Socio-Demographic Variable on the Sum Score Measures

Predictor	National Institutions		Local Institutions		Generalised	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Age	−0.03	0.01	−0.04	0.01	−0.03	0.02
Gender (ref: Female)	−0.05*	0.32	−0.06*	0.12	−0.06**	0.41
High Education (ref: Secondary)	−0.06**	0.50	−0.05	0.18	−0.07**	0.64
Low Education (ref: Secondary)	0.08**	0.40	0.03	0.15	0.07**	0.52
Ethnicity (ref: Minority)	0.36***	0.33	0.25***	0.12	0.35***	0.43
R-Squared	0.16		0.08		0.16	

* significant at 10%. ** significant at 5%. *** Significant at 1%.

Gender. Gender is associated with institutional trust based on each of the three summary measures. In each case, males are less trusting of the institutions.

Education. The level of education is associated with national and generalised institutional trust but not with trust in local institutions. In particular, education appears to have a negative linear effect on the two measures. As such, national and generalised institutional trust decrease as education increases.

Ethnicity. Higher trust in institutions is associated with the majority ethnicity for each of the three sum score measures. This is consistent with the findings from the analysis of individual items and the same interpretation is appropriate.

Factor Analysis without RSs Controlled. The initial one-factor model for trust in institutions fits the data poorly (see Table 6.4). A very large modification index with accompanying large expected parameter change (MI=198.65, EPC=0.35, standardised EPC=0.83) is observed for the error covariance between trust in the Mayor's/NDC office and trust in the RDC. These are the only two items that refer to local institutions and they form a separate dimension of institutional trust (also confirmed by a separate exploratory factor analysis with Mplus: RMSEA=0.07, CFI=0.95, TLI=0.90, SRMR=0.03). This contradicts the notion that institutional trust is unidimensional. As such the use of a sum score over all the items for generalised institutional trust in the Guyanese context is not appropriate.

The two-factor model (Two-Factor1) fits well overall with respect to the global fit indices (Table 6.4). However, further checks for misspecification using Jrule for Mplus 0.91 (Oberski, 2008) results in two freed error covariances: between army and national elections (MI=26.82, EPC=-0.16, Power=0.89) and between national elections and political parties (MI=16.15, EPC=0.11, Power=0.96). The negative covariance between the item for army and that for

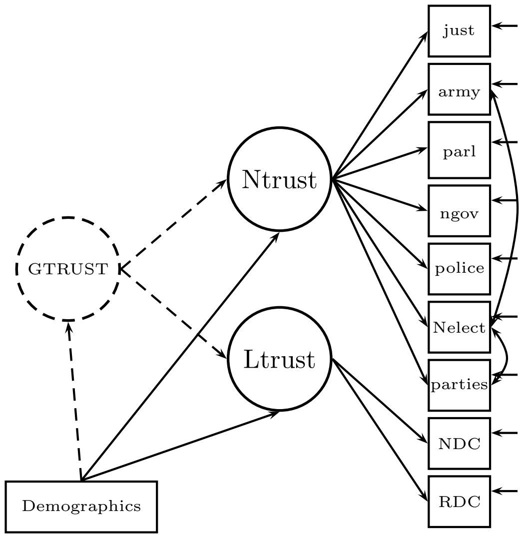

Figure 6.1: Factor Model for Institutional Trust Without Corrections for RSs

national elections indicate that citizens with higher trust in the national elections are less trusting of the army. This may be interpreted as that those who trust the elections more are less favourably disposed to supporting a military coup (Two-Factor2). The second freed error covariance is also plausible (Two-Factor3). Those who trust the national elections more are also more trusting of the political parties which compete at these elections. With these revisions, the model (Two-Factor3) fits the data adequately and no interpretable misspecifications remain (see Figure 6.1). In the analysis of the sum scores, these error covariances along with the error variances of each of the items could not be included. In this revised model, the convergent validity of each factor is adequate since the AVE is larger than 0.50 (Table 6.5). The two factors also display discriminant validity since the correlation among them (0.76) is smaller than the square root of the AVE for the local institutional trust (0.91) (Fornell & Larcker, 1981).

When a second-order factor is imposed on the two first-order factors, the fit of the model does not change. We refer to this second-order factor as generalised trust in institutions. However, we the interpretation of this measure is different from the generalised institutional trust based on the sum score

Table 6.4: Fit of the Factor Models

Model	χ^2	df	RMSEA	CFI	TLI	SRMR
Response Styles not Controlled						
One-factor	789.08	27	0.11	0.84	0.79	0.06
Two-Factor1	307.92	26	0.06	0.95	0.92	0.04
Two-Factor2	239.43	25	0.06	0.96	0.94	0.04
Two-Factor3	201.82	24	0.05	0.97	0.95	0.03
Second-Order	201.82	24	0.05	0.97	0.95	0.03
Structural Two-Factor	379.59	59	0.05	0.95	0.94	0.03
Structural Second-Order	397.64	64	0.05	0.94	0.93	0.04
Response Styles Controlled						
Two-Factor3	561.49	117	0.04	0.96	0.95	0.04
Second-Order	561.49	117	0.04	0.96	0.95	0.04
Structural Two-Factor	954.49	197	0.05	0.93	0.92	0.05
Structural Second-Order	972.52	202	0.05	0.93	0.92	0.06

because it correctly takes into account the sub-dimensions.

Two SEMs are estimated to provide the results in this section of the analysis. The first focuses on the first-order institutional trust factors (national and local). In this case, the second-order factor (generalised institutional trust) is not included in the model. The second SEM focuses on generalised institutional trust and here, the respondents' characteristics are allowed to impact only on the second-order factor (see Figure 6.1). Inclusion of the socio-demographic variables as predictors of the latent trust variables, does not affect the fit of the respective models substantially (See Table 6.4: Structural Two-Factor and Structural Second-Order). These variables jointly explain somewhat larger proportions of the variances in institutional trust (Table 6.6).

Age. Age has not effect on any of the institutional trust factors. This is consistent with the findings based on the sum scores.

Gender. Gender lacks any effect on institutional trust. This is in conflict with the results for gender when the sum scores are used as well as with some of the results from the individual-items analysis.

Education. The results for education are consistent with the results from the sum score models in indicating that education has a negative linear rela-

Table 6.5: Standardised Factor Loadings

Institution	RSs not Controlled		RSs Controlled	
Justice System	0.67		0.61	
Army	0.62		0.57	
Parliament	0.78		0.73	
National Government	0.82		0.77	
Police	0.70		0.63	
National Elections	0.73		0.67	
Political Parties	0.71		0.64	
Mayor's/ NDC Office		0.92		0.88
RDC Office		0.89		0.85
AVE	0.52	0.82	0.44	0.75

tionship with national and generalised institutional trust. However, the factor model indicates that this negative linear relationship is also relevant to trust in local institutions. Consequently, even if the items are correctly parcelled to provide the sum score measures, consistent results for the regression relationships between sum scores and factor models are not guaranteed. The effects of education in the factor model also reflect the findings from the individual-items analysis in only two out of nine cases.

Ethnicity. The results for both the sum sores and the individual-items measures indicate that the majority ethnic group is associated with higher trust in institutions. This general trend is confirmed by the factor model, but the coefficient for ethnicity is larger than in the sum score models.

Factor Analysis with RSs Controlled. The first step in the process of correcting for the RSs is establishing the RSs factors that will be included. The initial four-factor RIRSMACS model (containing ARS, ERS, DARS and MRS) fits the data adequately (RMSEA=0.03, CFI=1.00, TLI=1.00 and SRMR=0.03) and with the exception of DARS (AVE=0.37), each of the factors have adequate convergent validity with standardised loadings that range from 0.74 to 0.95. However, ARS lacks discriminant validity ($\sqrt{AVE}$=0.78, correlation with ERS=0.81). In addition, when the four RSs are included along with the institutional trust factors, the model fails to converge. We there-

Table 6.6: Predictors of Trust in Institutions with Trust Measured by Factor Models

Independent	Dependent					
	National Institutions		Local Institutions		Generalised	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
RSs not Controlled						
Age	−0.03	0.05	−0.03	0.05	−0.04	0.05
Gender (ref: female)	−0.05	0.04	−0.05	0.04	−0.06	0.04
High Education (ref: Secondary)	−0.07**	0.04	−0.07**	0.04	−0.07**	0.04
Low Eduucation (ref: Secondary)	0.08**	0.04	0.08**	0.04	0.07*	0.04
Ethnicity (fef: Minority)	0.39***	0.04	0.39***	0.04	0.39***	0.04
R-Square	0.19		0.09		0.19	
RSs Controlled						
Age	−0.03	0.05	−0.05	0.05	−0.04	0.05
Gender (ref: female)	−0.05	0.04	−0.06	0.04	−0.06	0.04
High Education (ref: Secondary)	−0.06	0.04	−0.04	0.04	−0.06	0.04
Low Education (ref: Secondary)	0.08**	0.04	0.04	0.05	0.07*	0.04
Ethnicity (ref: Minority)	0.37***	0.04	0.24***	0.05	0.37***	0.05
R-Squared	0.17		0.07		0.17	0.04

* significant at 10%. ** significant at 5%. *** Significant at 1%. The effects on national and local institutions are evaluated in the two-factor model whereas the effects on generalised trust is evaluated in the second-order model.

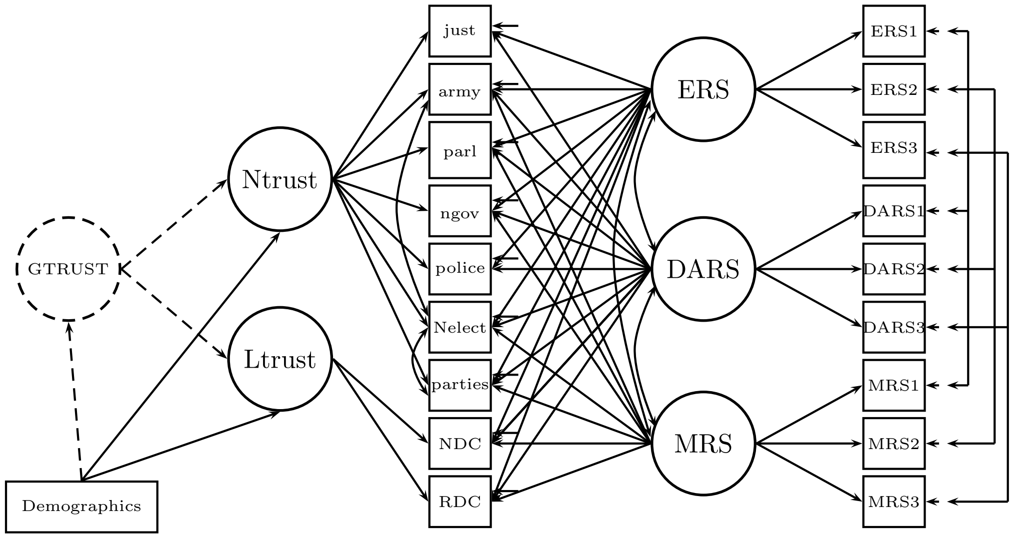

Figure 6.2: Factor Model for Institutional Trust with Corrections for RSs

fore, drop ARS from the model and proceed with only ERS, DARS and MRS. This revised three-factor RIRSMACS model fits adequately (RMSEA=0.04, CFI=1.00, TLI=0.98, SRMR=0.02). The convergent validity of the DARS factor (AVE=0.38) remains disappointing, but the factor loadings (0.63, 0.63, 0.59) are still relatively large.

To obtain the results for the factor model with the RSs controlled, we begin with the accepted factor model (Two-Factor3) (Table 6.2). The fit of this model is similar to that of the model in which the RSs are not controlled, but there are reductions in the standardised factor loadings (Table 6.5). The RSs therefore inflate the validity of the indicators and once they are controlled, the AVE for national institutional trust falls below 0.50. In spite of this, the factor loadings are still usefully large and the constructs continue to show adequate discriminant validity (correlation=0.72, $\sqrt{AVE}$ (local)=0.87).

Inclusion of the respondents' socio-demographic variables as predictors of institutional trust in the factor model with the RSs controlled alters the fit of the models, but there is still evidence of adequate fit (Table 6.4). As in the case where the RSs are not controlled, two separate prediction models are estimated: one containing no second-order factor and one containing the second-order factor with effects of the respondents' characteristics on only the second-order factor (see Figure 6.2). These SEMs explain approximately 17%, 7% and 17% of the variance in national, local and generalised institutional trust respectively (Table 6.6). These values are lower than when the RSs are not controlled, but the differences are not large. However, there are some differences in the effects of the respondents' socio-demographic characteristics in comparison to both the sum score models and the factor models in which the RSs are not controlled.

Age. The results for age are the same as for the factor model without RSs and for the sum score models. Age does not predict trust in institutions in Guyana.

Gender. Consistent with the factor model without the RSs controlled, gender does not explain institutional trust. This is in conflict with the results from the sum score models which indicate that males have lower trust in institutions.

Education. The findings about education when the RSs are controlled are in some cases different from both the sum score models and the factor models in which the RSs are not controlled. On the one hand, education does not have a negative linear effect on national and generalised institutional trust as observed in both the factor models without RSs and the sum score models. While the low education group has higher national and generalised institutional trust, the high education group does not have lower trust than the group with secondary education. The RSs therefore result in the significant effect observed in the previous factor model. On the other hand, when the RSs are controlled education does not predict local institutional trust. This is again in conflict with the findings from the factor model in which the RSs are not controlled. Interestingly, this lack of effect is consistent with the sum score model. Although factor models (and SEMs) are generally superior to sum score models (Bollen & Lennox, 1991; Shevlin et al., 1997), RSs can lead to spurious structural relationships and are hence alternative explanations for significant regression effects in institutional trust research.

Ethnicity. The majority ethnicity continues to have higher institutional trust than the other ethnicities when the RSs are controlled. This finding is consistent across all the models estimated which reinforces the conclusion that party ideology plays an important role in institutional trust in Guyana. Even though the effect of ethnicity remains significant when the RSs are controlled, there is a large drop in the size of the coefficient for its effect on trust in local institutions; from 0.39 to 0.24. The RSs therefore resulted in an approximately 63% increase in the size of the standardised effect. Apart from resulting in spurious effects altogether as observed for education, RSs can also distort the sizes of structural relationships.

6.7 Discussion

Although it is possible to measure generalised institutional trust in Guyana, it is not a first-order factor. The assumption that citizens of less consolidated democracies lack enough political sophistication to differentiate among categories of institutions (Mishler & Rose, 1997, 2005) seems to be too wide a generalisation. Based on our findings, we advise researchers against lumping the institutional trust items together to arrive at a single index without empirical investigation. The VAPO Guyana data set, includes items that cover both national and local political institutions and the respondents are able to differentiate between these different categories of institutions (Bouckaert & Van De Walle, 2001). Some researchers ignore variations in trust across institutions by using use a single item to measure trust in government. Our results indicate further, this approach may also ignore a differentiated view of institutional trust which may be relevant even in less advanced democracies. At the other end of the spectrum, many researchers analyse several items individually. With this approach, the results can quickly become too much to condense into meaningful summaries. Something between these two extremes is required.

Researchers should at least measure trust across several institutions and attempt to summarise these measures meaningfully. In this regard, the commonplace use of sum scores without verification (Poznyak et al., 2013) is not justified. In the analysis, we started with by assuming that institutional trust is unidimensional and can be summarised by a single sum. However, the factor models provide strong indications that this assumption is untenable. The population distinguishes between national and local institutions and this challenges the meaningfulness of the sum score approach. Research should be done to inform which items are combined to provide meaningful sum scores for the results to be meaningful, but even when the items are combined correctly into an index, the results may still be inaccurate (Neale et al., 2005).

Factor models are known to produce more accurate results than sum scores and the results remain accurate even when the reliabilities of the items are not high (Bollen & Lennox, 1991; Shevlin et al., 1997). Differences between the detected regression effects when SEM is employed compared to sum scores are therefore regarded as due to measurement errors (Neale et al., 2005; Shevlin et al., 1997). In the analysis, we find that both before and after the RSs are controlled in the factor models, gender is not a predictor of institutional trust, but when the sum score model is used, males are found to be less trusting. Sum scores can therefore lead to spurious regression effects. As a result, it is easy to understand what may occur in more complex explanatory models. Sum scores can distort all the regression relationships and it is therefore appropriate to question the accuracy of the conclusions.

Although factor analysis overcomes the outlined limitations of individual items and sum scores, it is not a panacea.[4] Even when factor models are used, RSs may still bias research results (Moors, 2012). In this study, we find that

[4]Several model based techniques inclusive of classification techniques may be applied to studying institutional trust.

the RSs inflate the factor loadings and convergent validity (Kankaraš & Moors, 2011; Welkenhuysen-Gybels et al., 2003). We also find that a negative linear relationship between education and institutional trust is due to the biasing effects of RSs. Only the result for the lowest education category is confirmed. Furthermore, an effect of education on trust in local institutions is altogether spurious and due to the RSs. Much of the literature indicates that education has a negative relationship with institutional trust (Blanco, 2013; Hutchison & Johnson, 2011; Lühiste, 2006; Rohrschneider & Schmitt-Beck, 2002). The current findings suggest that this may be due to either the use of sum scores or to RSs. While factor models for institutional trust improve the results in comparison to sum scores, RSs may still bias the findings and are alternative explanations for the significant regression relationships detected. As such RSs must be controlled in institutional trust research.

One of the challenges that researchers will face in correcting for RSs stems from the fact that they often use exploratory factor analysis. While, several methods of correcting for RSs are available (see Van Vaerenbergh & Thomas, 2013, for a review), they do not apply to this technique. One possibility is that researchers may adjust the individual items for RSs before using this approach. This may be done by regressing each item on the RSs then using the residual as the new variable (Baumgartner & Steenkamp, 2001). An important limitation of this approach is that the measures of the RSs themselves contain measurement error which gets passed on to the corrected scores due to the linear regression of the items on the RSs, thereby introducing new errors (Weijters et al., 2008). Researchers are therefore advised to use the CFA framework to analyse their data instead of exploratory factor analysis as this allows appropriate corrections for the RSs. The RSs model implemented in this study requires separate heterogeneous items (uncorrelated and measuring different constructs) to ensure that content is not confounded with style. To correct for the RSs with the RIRSMACS model at least 6 heterogeneous items are required, but 15 such items is recommended. Note that more items are required for a dedicated RSs study (Weijters et al., 2008). However, there are other methods of correcting for RSs that have different requirements and researchers can choose methods depending of their research constraints, but it is important to understand the benefits and limitations of the methods before making a selection (see Van Vaerenbergh & Thomas, 2013, for a review of the methods).

In future work, researchers should use factor models with corrections for RSs instead of individual items or sum scores. It is also necessary for the established relationships between institutional trust and other variables to be re-examined with the RSs controlled in light of their adverse effects. The differentiated view of national and local institutions in Guyana along with the variations across several single items presents a challenge to the practice of using only a single item to measure trust in government. To determine the extent to which such a measure is appropriate, attempts should be made to correlate the generalised institutional trust factor with the single–item measure of trust in government.

References

Abts, K. (2012). *Maatschappelijk onbehagen en etnopopulisme: Laatmoderniteit, burgers, ressentiment, migranten, politiek en extreemrechts* (Unpublished doctoral dissertation). Faculty of Social Sciences, University of Leuven.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Bean, C. (2003). Citizen confidence in social and political institutions in a changing world. In C. Bailey, D. Cabrera, & L. Buys (Eds.), *Social change in the 21st century conference.* Brisbane: Centre for Social Change Research, Queensland University of Technology.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Blanco, L. R. (2013). The impact of crime on trust in institutions in Mexico. *European Journal of Political Economy*, *32*, 38–55.

Blanco, L. R., & Ruiz, I. (2013). The impact of crime and insecurity on trust in democracy and institutions. *American Economic Review*, *103*(3), 284–288.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Bouckaert, G., & Van De Walle, S. (2001). Government performance and trust in government. Paper presented at the European Group for Public Administration Annual Conference, 5–8 September, Vaasa, Finland.

Byrne, B. (2012). *Structural equation modeling with mplus: Basic concepts, applications and programming.* New York: Taylor & Francis Group.

Chang, E. C. C., & Chu, Y.-h. (2006). Corruption and trust: Exceptionalism in Asian democracies? *The Journal of Politics*, *68*(2), 259–271.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Christensen, T. O. M., & Læ greid, P. (2005). Trust in government: The relative importance of service satisfaction, political factors, and demography. *Public Performance & Management Review*, *28*(4), 487–511.

Fornell, C., & Larcker, D. F. (1981). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Journal of Marketing Research*, *18*, 39–50.

Hamm, J. A., PytlikZillig, L. M., Tomkins, A. J., Herian, M. N., Bornstein, B. H., & Neeley, E. M. (2011). Exploring separable components of institutional confidence. *Behavioral sciences & the Law*, *29*(1), 95–115.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Huang, K.-p., Lee, F.-y., & Lin, T.-m. (2013). Partisanship and institutional trust a comparative analysis of emerging democracies in East Asia. *Taiwan Journal of Democracy*, *9*(1), 47–71.

Hug, S., & Spörri, F. (2011). Referendums, trust, and tax evasion. *European Journal of Political Economy*, *27*(1), 120–131.

Hutchison, M. L., & Johnson, K. (2011). Capacity to trust? institutional capacity, conflict, and political trust in Africa, 2000-2005. *Journal of Peace Research*, *48*(6), 737–752.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Khodyakov, D. (2007). Trust as a process: A three-dimensional approach. *Sociology*, *41*(1), 115–132.

Listhaug, O. (1984). Confidence in institutions: Findings from the Norwegian Values Study. *Acta Sociologica*, *27*(2), 111–122.

Lühiste, K. (2006). Explaining trust in political institutions: Some illustrations from the Baltic states. *Communist and Post-Communist Studies*, *39*(4), 475–496.

Mishler, W., & Rose, R. (1997). Trust , distrust and skepticism: Popular evaluations of civil and political institutions societies in post-communist societies. *The Journal of Politics*, *59*(2), 418–451.

Mishler, W., & Rose, R. (2001). What are the origins of political trust?: Testing institutional and cultural theories in post-communist societies. *Comparative Political Studies*, *34*(1), 30–62.

Mishler, W., & Rose, R. (2005). What are the political consequences of trust?: A test of cultural and institutional theories in Russia. *Comparative Political Studies*, *38*(9), 1050–1078.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Neale, M. C., Lubke, G., Aggen, S. H., & Dolan, C. V. (2005). Problems with using sum scores for estimating variance components: contamination and measurement noninvariance. *Twin research and human genetics: The official journal of the International Society for Twin Studies*, *8*(6), 553–68.

Oberski, D. (2008). Jrule for Mplus. Retrieved from `https://github.com/daob/JruleMplus/wiki`

Poznyak, D., Meuleman, B., Abts, K., & Bishop, G. F. (2013). Trust in american government: Longitudinal measurement equivalence in the ANES, 19642008. *Social Indicators Research*.

Rohrschneider, R., & Schmitt-Beck, R. (2002). Trust in democratic institutions in Germany: Theory and evidence ten years after unification. *German Politics*, *11*(3), 35–58.

Rothstein, B., & Stolle, D. (2008). The state and social capital: An institutional theory of generalized trust. *Comparative Politics*, *40*(4), 441–459.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582.

Secor, A. J., & Loughlin, J. O. (2005). Social and political trust in Istanbul and Moscow: A comparative analysis of individual and neighbourhood effects. *Transactions of the Institute of British Geographers*, *30*(1), 66–82.

Shevlin, M., Miles, J., & Bunting, B. (1997). Summated rating scales. a monte carlo investigation of the effects of reliability and collinearity in regression models. *Personality and Individual Differences*, *23*(4), 665–676.

Shlapentokh, V. (2006). Trust in public institutions in Russia: The lowest in the world. *Communist and Post-Communist Studies*, *39*(2), 153–174.

Thomas, T. D., Abts, K., & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement*, *74*(1), 97–115.

Thomas, T. D., Abts, K., & Vander Weyden, P. (in press). Measurement invariance, response styles and rural-urban measurement comparability. *Journal of Cross-Cultural Psychology*.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies.

International Journal of Public Opinion Research, *25*(2), 195–217.

Van der Veld, W. M. (2008). Judging misspecifications in structural equation models using JRule. *Symposium on Testing Structural Equation Models at the ISA-RC33, 7th International Conference on Social Science Methodology. Naples (Italy), September 1–5*, 1–5.

Van der Veld, W. M., & Saris, W. E. (2011). Cause of generalized social trust. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 207–247). New York: Routledge: Taylor & Francis Group.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana.* Ghent, Belgium: Ghent University.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*, 29–37.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

CHAPTER 7

A COMPARISON OF SALIENT RESPONSE STYLES BETWEEN LATENT CLASS ANALYSIS AND CONFIRMATORY FACTOR ANALYSIS

A Comparison of Salient Response Styles between Latent Class Analysis and Confirmatory Factor Analysis

Troy Devon Thomas[1,2], Koen Abts[3] and Patrick Vander Weyden[4]

[1]*Department of Comparative Science of Cultures, Ghent University.*

[2]*Department of Mathematics, Physics and Statistics, University of Guyana.*

[3]*Centre for Sociological Research, University of Leuven.*

[4]*Founder of the independent research company — FocusUP.*

Abstract

This paper compares the salient response styles (RSs) in the Guyanese population between latent class analysis (LCA) and confirmatory factor. The LCA model detects extreme RS and two milder styles that are often overlooked in CFA models, but both of which describe larger proportions of the population than does ERS. It is therefore likely that researchers who use CFA are systematically modelling and correcting for less salient RSs in various populations. As such, a modified approach to determining the RSs to include in CFA models is proposed. There is also high convergent validity of the salient RSs typologies between LCA and CFA, but the effects of the respondents' characteristics are not entirely consistent. These results are based on the representative indicators measures of the RSs. In addition to discussing these issues, this paper provides guidelines on how to correct for RSs with representative indicators in LCA models.

Keywords: response styles, latent class analysis, confirmatory factor analysis, response styles predictors, convergent validity, measurement

---

7.1 Introduction

Response styles (RSs) which are the respondents' systematic tendency to respond in certain ways to rating scale items regardless of content, bias research results (Baumgartner & Steenkamp, 2001; Thomas, Abts, & Vander Weyden, in press). Several model-based methods have been developed for RSs and these often require latent class analysis (LCA) (Moors, 2003), or confirmatory factor analysis (CFA) (Billiet & McClendon, 2000; Weijters, Schillewaert, & Geuens, 2008). However, the particular RSs modelled with LCA are due to their salience in the data (Moors, 2003, 2012) whereas established research practice determines the styles included in CFA models. As such, different RSs might be modelled with the two techniques even when the same data are analysed.

CFA models are most likely to include some combination of acquiescence RS (ARS: tendency to agree), extreme RS (ERS; tendency to use scale endpoints), midpoint RS (tendency to select scale midpoint) and disacquiescence RS (DARS: tendency to disagree). These are the traditionally more recognised RSs. LCA models, on the other hand, often do not include ARS, MRS and DARS, but may include ERS and mild RS (MLRS: tendency to avoid scale endpoints). Given that the RSs modelled with LCA are based on their salience in the data, we question whether CFA models include the important styles given the population under study.

Besides, the input information and the assumptions of the two modelling techniques are different (Wang & Xiaoqian, 2012) and this may contribute to different results. In particular, conflicting results seem to occur for the effect of the respondents' variables on RSs in separate studies (Van Vaerenbergh & Thomas, 2013). Such conflicts may imply a lack of convergent validity which overshadows the entire endeavour of modelling and correcting for RSs. In spite of this, we encounter no studies that directly investigate the convergent validity of RSs between LCA and CFA.

To assist in filling the gaps in the literature, this paper compares the results for RSs between LCA and CFA with survey data from Guyana. It: (1) identifies the salient RSs typologies (LCA) in the population and proposes an approach to determining which RSs to include in CFA models, (2) investigates the convergent validity of the detected RSs between LCA and CFA, and (3) compares the relationships between the detected RSs and the respondents' characteristics between the two techniques. These issues are investigated using representative indicators to measure the RSs.

7.2 Latent Class Analysis and Factor Analysis

Although there are several variants of LCA, LCA models with ordered categorical or nominal (categorical) latent variables is often employed in research. Such categorical LCA is the subject of this paper. LCA focuses on categorising individuals by identifying sub-populations consisting of similar cases (Wang & Xiaoqian, 2012). In contrast, CFA is variable-centred. It focuses on the relationships among variables and regards the latent factors as continuous (Moors, 2003). In addition, LCA takes the full cross-classification tables of the variables as input whereas CFA takes the variance-covariance matrix of the manifest variables (McCutcheon, 1987; Wang & Xiaoqian, 2012).

LCA and CFA are analogous, but there are some differences. LCA derives classes in such a way that the indicators are locally independent. The assignment of individuals to one of finitely many classes is probabilistic as are the response patterns to the set of indicators (Hagenaars & McCutcheon, 2002). In contrast, the manifest variables are regarded as imperfect measures of the factors in CFA and the factor loadings indicate the strength of association between the manifest variables and the latent factor. The overarching concept is that the factor explains the correlations among the items (Wang & Xiaoqian, 2012).

Both LCA and CFA have been further developed to accommodate indicators at various measurement levels (Kankaraš & Moors, 2009), but the techniques differ with respect to the assumptions made. CFA requires multivariate normality and homogeneity of the population distribution, but LCA does not. Furthermore, LCA can handle non-linear relationships since it models the responses to each category of the indicators whereas CFA assumes a linear monotonic relationship between the factors and the indicators (Moors, 2012).

LCA and CFA may also produce different results for the same data. For example, LCA is fairly accurate at identifying metric and scalar invariance, but CFA may incorrectly indicate metric non-invariance instead of scalar non-invariance (Kankaraš, Vermunt, & Moors, 2011). If modelling technique dependency occurs in RSs models, further errors may actually be introduced by attempting to correct for RSs. It is therefore also appropriate to question whether or not relationships between the RSs and the respondent characteristics depend on the estimation technique employed.

7.3 Response Styles Models

RSs are traditionally approached from two perspectives. They are viewed as either nuisances which should be controlled while focusing on other topics, or as meaningful personality constructs which should be studied. This means that they are regarded as either situational- or person-dependent respectively (Van Vaerenbergh & Thomas, 2013). Both perspectives are supported: there is a large time-invariant component of the RSs, which is attributed to the respondents (Billiet & Davidov, 2008; Weijters, Geuens, & Schillewaert, 2010), but there is also evidence of situational effects since some respondents tend to switch between ERS and MLRS over time (Aichholzer, 2013). In general, the occurrence of situational effects does not invalidate the view of RSs as respondent traits since situational variables may encourage or discourage inherent tendencies of individuals (Baumgartner & Steenkamp, 2001).

Corresponding to the two major perspectives, researchers have proposed methods that either seek only to correct for RSs or methods that permit studying the RSs themselves. One example of correcting for an RS is the use of balanced scales (for ARS) (Cloud & Vaughn, 1970). Such scales are thought to result in a cancelling-out of ARS because they include both negatively and positively worded items targeting the same issue (Baumgartner & Steenkamp, 2001). In contrast, the representative indicators approach which uses separate, heterogeneous items to estimate the RSs is used to both study and correct for the RSs (Greenleaf, 1992a, 1992b). This approach therefore combines the two seemingly divergent perspectives on RSs.

More recently, researchers have transformed many of the basic RSs measurement and correction schemes into statistical models. These are most often implemented using LCA, CFA and Item Response Theory (IRT). IRT models are used mainly for ERS (Jin & Wang, 2014) whereas LCA and CFA are often used to model several RSs simultaneously. Both LCA and CFA facilitate RSs estimation with or without balanced scales (style factor) and with representative indicators.

The style factor approach to modelling RSs, involves specifying an RS factor(s) on the items measuring substantive constructs. The CFA style factor (method factor) requires balanced scales and it models ARS (Billiet & Davidov, 2008; Billiet & McClendon, 2000; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). When LCA is employed, the scales may or may not be balanced and several RSs may be modelled simultaneously (Moors, 2003, 2012). For example, researchers have successfully modelled ERS (Moors, 2003, 2004), ERS and ARS (Kieruj & Moors, 2013; Moors, 2012) and MRS (Moors, 2008) with LCA. Comparisons between CFA and LCA based on style factors will therefore be restricted to ARS. Furthermore, the exploratory way in which the LCA style factors are identified[1] may also result in ARS not being included at

[1]LCA may also be confirmatory. However, the specific RSs modelled is determined from the results.

all (Moors, 2003); thus negating comparison.

Although representative indicators approaches to modelling RSs may be implemented with both LCA and CFA (for example, Aichholzer, 2013; Weijters et al., 2008), adjustments for RSs with LCA using representative indicators have not been demonstrated as yet. Aichholzer uses representative indicators within the LCA framework to study ERS and MLRS. However, it is possible to have other RSs included in such models. Within the CFA framework, the Representative Indicators Response Styles Means and Covariance Structure (RIRSMACS) model uses dedicated RSs items to model ARS, DARS, ERS and MRS and it allows for studying and controlling RSs (Weijters et al., 2008). This model is also flexible enough to include other RSs (for example, Weijters, Geuens, & Schillewaert, 2010). Representative indicators within the LCA framework and the RIRSMACS model within the CFA framework provide an opportunity to compare LCA and CFA with respect to RSs.

7.4 Research Questions

The issues investigated in this paper may be summarised into three questions:

1. Is the apriori inclusion of some combination of ARS, ERS, DARS and MRS in CFA models optimal?

2. What is the extent of convergent validity of the salient RSs between LCA and CFA?

3. How consistent are the effects of the respondents' characteristics on the salient RSs between LCA and CFA?

These questions are discussed in the remainder of this section.

RSs Selection. LCA models detect the salient RSs (Kieruj & Moors, 2010, 2013; Moors, 2003, 2012). As such, there are no guarantees that an LCA RSs model will include some combination of ARS, ERS, DARS and MRS as is usual the RIRSMACS (CFA) model. Therefore, a comparison of the LCA results to these four popular RSs enables evaluation of the CFA approach of deciding which styles to model beforehand.

Convergent Validity. Convergent validity between techniques is important in fostering confidence in RSs measurements and corrections. However, there are no studies examining the convergent validity of RSs between LCA and CFA. Furthermore, there is not much research on the convergent validity across different methods of measuring RSs in general (Van Vaerenbergh & Thomas, 2013).

Nevertheless, low convergent validity between the representative indicators approach (Greenleaf, 1992a, 1992b) and methods which do not control the content of the items leads to warnings against using adhoc RSs measurements (De Beuckelaer, Weijters, & Rutten, 2010). In addition, when a representative indicators measure of ERS (percentage of extreme responses) is correlated with an LCA ERS style factor, the results (correlations = 0.37 and 0.49) are more encouraging (Kieruj & Moors, 2013). Low convergent validity between the RSs modelled with LCA and CFA will indicate modelling technique specificity whereas high convergent validity bodes well for RSs measurement and correction across the techniques.

Respondents' Characteristics. The demographic characteristics of respondents are linked to RSs (Thomas, Abts, & Vander Weyden, 2014; Weijters, Geuens, & Schillewaert, 2010), but there are several inconsistencies in the nature of the relationships. The question is whether or not the inconsistencies are due to the modelling techniques.

Based on the style factor approach in the CFA framework, Billiet and McClendon (2000) indicates that age is positively whereas education is inversely related to ARS. In contrast, Kieruj and Moors (2013) find no effects of age, gender and education on ARS within the LCA framework. The results for ERS based on LCA models indicate consistently that gender and education have no effect (Kieruj & Moors, 2013; Moors, 2008), but they are contradictory in relation to the effect of age. Whereas Moors finds no effect of age, Kieruj and Moors report that older respondents are more likely to use ERS.

Between-study differences in the effects of the respondents' characteristics also occur with representative indicators. Using LCA, Aichholzer (2013) finds that extreme responders are less well-educated and older than mild responders and that gender is unrelated to ERS. Based on the RIRSMACS model, Weijters, Geuens, and Schillewaert (2010) find that: (1) age is positively related to ARS, ERS and MRS but not related to DARS, (2) ARS and ERS are higher among females but gender does not affect DARS and MRS, and (3) education is inversely related to ARS, ERS and MRS but not related to DARS. In contrast, Thomas et al. (2014) indicate that the effects of the respondents' variables vary across within-country, rural-urban subcultures depending on the RS. However, they report that age and gender do not affect DARS and MRS; gender does not predict ERS; education is inversely related to ARS; and that the majority ethnic group uses less ERS.

The described results highlight similarities and differences in the effects of the respondents' characteristics between the methods of calculating the RSs and between CFA and LCA. Studies using the same approach sometimes contradict each other and this may be due to differences in the populations. To avoid possible composition effects, we investigate the comparability of the results for the respondents' characteristics using a single data set. The same general method of measuring the RSs – representative indicators – is also used with both techniques.

7.5 Data and Methods

7.5.1 Data

The data are obtained from the Values and Poverty Study in Guyana (VAPO Guyana) which was conducted between April and May 2012. The study was funded by the Flemish Inter-University Counsel (VLIR) and jointly executed by the University of Guyana and Ghent University. It investigates both methodological and substantive issues and provides an opportunity to study RSs with representative indicators. The data were collected via face-to-face interviews by a survey organisation (DPMC) under the supervision of the Universities of Guyana and Ghent. The interviewers were trained by DPMC and they attended a two-day briefing session organised by the VAPO research team (Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012).

The VAPO Guyana employed a two-step sampling procedure which randomly selected municipalities with probability proportional to size, and respondents within the municipalities with equal probabilities. The procedure resulted in the selection of 87 clusters within 51 municipalities. In total, 1048 individuals were interviewed at a response rate of 87% (American Association for Public Opinion Research, 2011, RR2). The data are weighted by iterative proportional fitting and are representative of the coastal residents (specifically Region 2, 3, 4, 5, 6, and 10), who account for approximately 90% of the total population of the country (Bureau of Statistics, 2002).

The respondent characteristics included in this study are age, gender, education and ethnicity. Age is continuous and it is measured in years (average = 36.25). Gender is dichotomous: 1 = male (49%) and 0 = female (51%). Education represents the level of schooling completed and it is coded into three levels: 1 = primary education (31%); 2 = secondary education (57%); and 3 = above secondary education (13%). Ethnicity is dichotomous with 1 representing the largest group (46%) – East Indians – and 0 representing the other ethnicities (54%) - Afro, Amerindians, Chinese, Mixed, Portuguese, and White.

Before proceeding further, it is important to clarify an issue about the use of survey data in this study. This paper does not focus on information recovery. Hence, simulation data are not crucial. Once the salient RSs are identified, the convergent validity and consistency of predictions between the two techniques maybe evaluated. In this paper, re restrict to the salient RSs that emerge from the data.

7.5.2 Methods

The RSs are measured with representative indicators in both the LCA and the CFA models. Representative indicators control the content of the items by requiring that they are unrelated (Weijters et al., 2008). This study uses 27 heterogeneous attitude items (see Appendix A.2) with an average interitem correlation of 0.06. The scale format is also controlled. Although some indicate that the number of scale categories do not affect ERS with end-labelled scales (Kieruj & Moors, 2010, 2013), others find that the number of scale categories is inversely related to ERS and that ERS decreases with fully labelled scales (Weijters, Cabooter, & Schillewaert, 2010). Each item used to measure the RSs in this study is scored on a 5-point fully labelled rating scale. Hence, the scale format does not differentially affect the respondents' use of RSs between the groups.

In the analysis, the LCA model is implemented first to determine the salient RSs typologies. Subsequently, these RSs types are estimated with CFA; specifically the RIRSMACS model. This approach breaks the tradition of modelling ex-ante particular RSs with CFA and enables modelling the styles detected in the data. This ensures that the same RSs are modelled using LCA and CFA which facilitates comparisons of the techniques.

LCA Implementation. The LCA model is implemented with Mplus 7.11 (L. K. Muthén & Muthén, 1998–2012). Model selection is based on a combination of the AIC, BIC, adjusted BIC and the Mendel-Rubin Adjusted LRT in addition to the interpretability of the extracted RSs typologies (Kankaraš, Guy, & Vermunt, 2011; Nylund, Asparouhov, & Muthén, 2007).

To implement the model, the 27 RSs items are used as the outcome variables of a single categorical latent variable (see Figure 7.1). The classes of this latent variable are the RSs. The estimation begins with two classes and the number of classes is increased in a step-wise manner until best model is obtained.

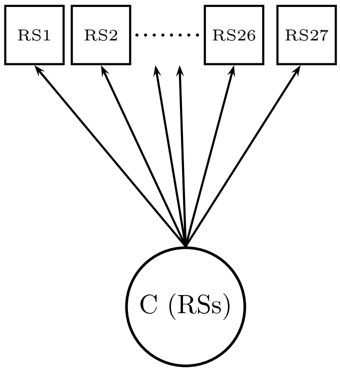

Figure 7.1: Latent Class RSs Model

The large number of items and the number of response categories would lead to sparseness in the classification tables. However, the model is simplified by imposing some constraints. In each latent class, all the respective item thresholds are equated, but they are allowed to vary between the classes. Consequently, one set of thresholds (4 thresholds for the 5-category scales) and one set of probabilities (5 – one per response option) of choosing a particular response option given class membership per latent class is estimated for all 27 items. These equality constraints are necessary within the representative indicators paradigm in which the effect of each RS is constant across heterogeneous items (Moors, 2012).

Following identification of the salient RSs, the effects of the respondents' variables are evaluated. These are estimated with a multinomial logit (MLg) model in which the latent classification (RSs) is the dependent variable (Wang & Xiaoqian, 2012).

CFA Implementation. The CFA model implemented is the RIRSMACS model (Weijters et al., 2008) and it is implemented with Mplus 7.11. For this model, the 27 RSs items are divided at random into three blocks of 9 items each and one indicator per RS is calculated from each block. Therefore, each RS has three indicators. The calculation of the values of the indicators are based on the method used by Weijters et al. (2008), but we wait until the included styles are determined to describe the calculations. An example of a RIRSMACS model with four RSs is shown in Figure 7.2.

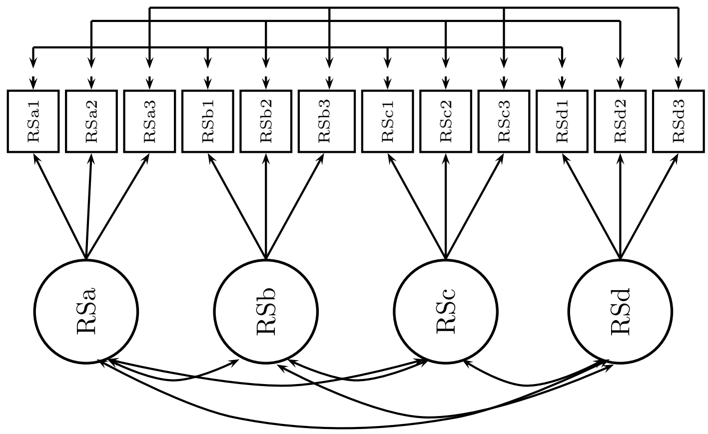

Figure 7.2: A RIRSMACS Model with Four Response Styles

In the model, the RSs factors are allowed to correlate freely and the error terms of the indicators calculated from the same block are correlated.[2] The first factor loading for each factor is equated to 1 to scale the factors, but they are no other constraints on the sizes the factor loadings. In addition, cross-loadings are not allowed in the model (See Figure 7.2).

Before estimating the RIRSMACS model, a decision has to be made on which RSs to include. Furthermore, the RSs modelled must be the same between the LCA and the RIRSMACS model to permit comparisons. To facilitate this, the CFA model is implemented subsequently to the LCA model and the results of the LCA implementation is used to ensure that the same RSs are estimated with CFA. As noted earlier, the RIRSMACS model is flexible enough to accommodate the RSs identified.

To evaluate the convergent validity of the RSs between LCA and CFA, the latent class assignments are used to predict the RIRSMACS, RSs factors

[2]The fit of the CFA and subsequent LISREL model which includes the respondent variables as predictors of the RSs is evaluated using the alternative fit indices: Root Mean Square Error of Approximation (RMSEA) less than or equal to 0.06, Comparative Fit Index (CFI) and Tucker and Lewis Index (TLI) greater than 0.95 and Standardised Root Mean Square Residual (SRMR) less than or equal to 0.05, indicate acceptable overall fit (B. Byrne, 2012; B. M. Byrne, Shavelson, & Muthen, 1989; Hu & Bentler, 1999).

using a LISREL model. Convergent validity is determined from the proportion of the variances (and correlations) of the RSs factors explained by the class assignments. Following this, the effect of the respondents' characteristics on the RIRSMACS factors are examined with a LISREL model. This facilitates comparisons with the results of the MLg model (LCA results).

7.6 Results

7.6.1 Typologies of RSs

The fit of the LCA model improves with more classes based on the AIC, BIC and ABIC values (Table 7.1) and up to five classes with respect to the Mendel-Rubin Adjusted LRT (2 to 4 classes, p-value < 0.004; 5 classes, p-value $= 0.045$; 6 classes, p-value $= 0.127$). The quality of the class assignments is also high for each model (Table 7.1). However, we stop at the 8-class because no new RSs are identified. From the 7-class solution onwards, some refinements (duplicates) of the classes already identified emerge.

Up to and including the 4-class model, the conditional probabilities of the item categories highlight several distinct and interpretable RSs (Figures 7.3,7.4,7.5, and 7.6). The two-class model (Figure 7.3) identifies ERS and MLRS. The three-class solution adds a group of respondents who tend to avoid both the scale endpoints and midpoint (Figure 7.4). This style is not previously assessed in either LCA or CFA models, but it is approximately twice as popular as ERS. We refer to this style as mild directional RS (MDRS). The fourth class in the four-class model identifies those respondents who use no particular RS (see Figure 7.5). The emergence of this class is important since these ideal respondents would otherwise be lumped together largely with the mild responders. It is also interesting to note that this group of respondents is large in comparison to the groups that use one of the RSs.

A few difficulties emerge beginning with the 5-class model (Figure 7.6). The 5-class solution identifies what we label as ARS*. However, only the high-

Table 7.1: LCA Model Selection

Classes	AIC	BIC	ABIC	Class Proportions								Entropy
				C1	C2	C3	C4	C5	C6	C7	C8	
				ERS	MLRS	MDRS	No RS	ARS*	MRS*	MDRS2	MLRS2*	
2	75330.46	75375.05	75346.46	0.19	0.81							0.96
3	73448.08	73517.44	73472.98	0.14	0.60	0.26						0.92
4	72497.39	72591.53	72531.18	0.12	0.19	0.25	0.45					0.89
5	72118.11	72237.02	72160.80	0.06	0.13	0.23	0.45	0.13				0.89
6	71909.32	72053.01	71960.09	0.06	0.12	0.19	0.45	0.13	0.06			0.89
7	71742.38	71910.84	71802.85	0.06	0.13	0.14	0.29	0.13	0.06	0.19		0.84
8	71618.70	71811.93	71688.07	0.05	0.11	0.13	0.13	0.09	0.05	0.16	0.27	0.83

Class assignment is based on the most likely class membership. MDRS: Mild Directional RS.
* indicates that the RS is not unequivocally established.

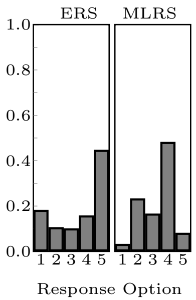

Figure 7.3: Probabilities for two-class solution

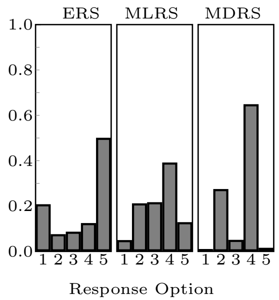

Figure 7.4: Probabilities for three-class solution

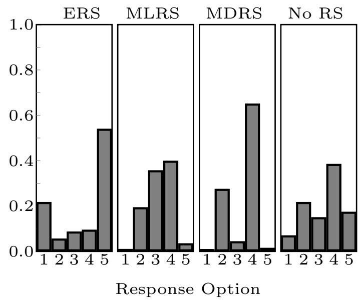

Figure 7.5: Probabilities for four-class solution

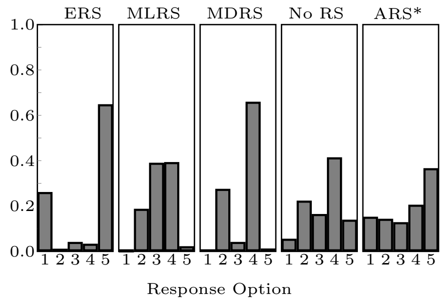

Figure 7.6: Probabilities for five-class solution

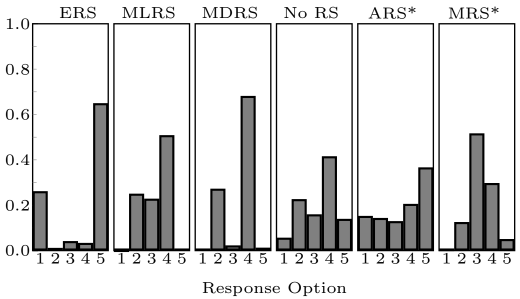

Figure 7.7: Probabilities for six-class solution

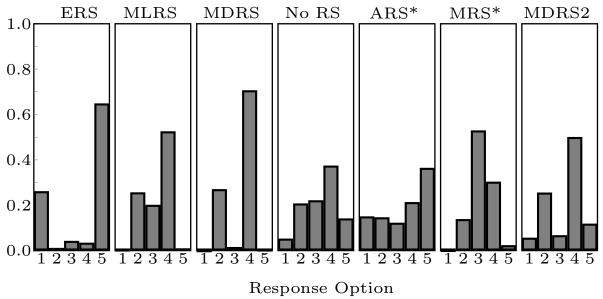

Figure 7.8: Probabilities for seven-class solution

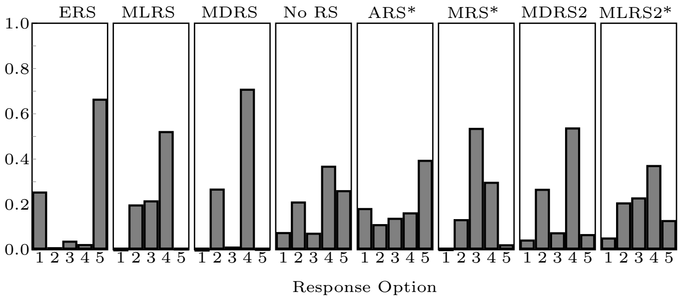

Figure 7.9: Probabilities for eight-class solution

est response category stands out from the middle and two lowest categories. The expected ordinal form of ARS with higher categories (stronger agreement) preferred over lower categories (Moors, 2012) is not well-demonstrated. Furthermore, the pattern becomes less convincing as the number of classes is increased. The 6-class model (Figure 7.7) adds a category which we label as MRS*. In this class, the probabilities for disagree and agree in particular are higher than expected if this styles were really MRS. As in the case of the 5-class solution, the evidence of the RSs (MRS here) is very weak.

Neither of the categories introduced by the 5- and 6-class solutions is formed well enough to be regarded as a salient RSs. The addition of the "*" to the names is meant to convey the reluctance with which we refer to them as ARS and MRS. We use these labels because they are the RSs to which these categories are most closely related. However, the names assigned are not meant to indicate that these categories truly represent ARS and MRS respectively.

The 7- and 8-class solutions (see Figures 7.8 and 7.9) add complexity to the model by including additional MDRS and MLRS (not strongly evidenced) categories. Although these categories results in a better-fitting model with respect to the AIC, BIC and adjusted BIC values, they are unnecessary and are disregarded. The 5 and 6-class solutions are similarly rejected because the categories formed cannot be clearly interpreted as RSs. Including these classes in an RSs correction model is likely to introduce errors rather than correct for RSs errors. As such, the 4-class model is accepted as the best fitting an interpretable RSs model.

7.6.2 CFA Implementation of Salient RSs

The finding that ARS, DARS and MRS do not describe salient subgroups of respondents is striking given the relative popularity of these styles in CFA models. MLRS and MDRS which seem to be important in the Guyanese population are often not considered. If researchers wish to model and correct for the salient RSs in a population using CFA, it is advisable that they determine the salient RSs from LCA studies done on the same group. There is therefore a need for RSs studies based on LCA to assist CFA researchers. Otherwise, researchers may follow an alternative stepwise approach. First, they can model the salient RSs with a classification technique such as LCA. Second, once the meaningful styles in a population are determined, CFA models should be modified to accommodate them.

For the CFA implementation in this study, the RIRSMACS model is modified to include ERS, MLRS and MDRS. The values of the RSs indicators are calculated as:

$$ERS = [f(1) + f(5)]/k$$

$$MLRS = [f(2) + f(3) + f(4)]/k$$

and

$$MDRS = [f(2) + f(4)]/k,$$

where $f(x)$ is the frequency of the response option x and $k = 9$ is the number of items per block (based on Weijters et al., 2008). These calculations are repeated on each block of items so that three indicators per RS are obtained.

The overall fit of the CFA model is excellent (RMSEA=0.02, CFI=1.00, TLI= 1.00, SRMR=0.02) and the item validity (factor loadings) and factor convergent validity (average variance extracted) are also high; greater than 0.50 (Table 7.2) (Fornell & Larcker, 1981). In the model, the ERS and MLRS factors have a strong negative correlation of -0.98 (see Table 7.3). Given that these two styles are opposites conceptually, the negative correlation is logical. This would introduce strong collinearity into the model. However, to correct for RSs with the RIRSMACS model, researcher need not include both ERS and MLRS since by controlling ERS for example, the variation in the items due to MLRS is already controlled. MLRS and MDRS also overlap since they share

Table 7.2: Standardised factor loadings in the CFA model

Item Code	Standardised Factor Loadings		
	ERS	MLRS	MDRS
ERS1	0.86		
ERS2	0.89		
ERS3	0.95		
MLRS1		0.85	
MLRS2		0.88	
MLRS3		0.95	
MDRS1			0.86
MDRS2			0.85
MDRS3			0.90
Average Variance Extracted	0.81	0.80	0.76

Note: 1, 2, and 3 in the item code identify the block from which the indicator was calculated.

Table 7.3: Correlations among the RIRSMACS factors

	ERS	MLRS	MDRS
ERS	**0.90**		
MLRS	-0.98	**0.89**	
MDRS	-0.84	0.86	**0.87**

The square root of the average variance extracted are on the diagonal and in bold font. The off-diagonal elements are the correlations between the respective pairs of factors.

all but the scale midpoint. As expected, MLRS lacks discriminant validity and it is removed form the CFA model.

The revised model fits adequately (RMSEA = 0.03, CFI = 1.00, TLI = 1.00, SRMR = 0.01) and there are small changes to the standardized factor loadings: ERS (0.87, 0.89, 0.94), MDRS (0.87, 0.90, 0.85). In spite of these changes, the convergent and discriminant validity of ERS and MDRS are unchanged.

7.6.3 Convergent Validity of the RSs between LCA and CFA

To evaluate the convergent validity of the RSs across the two techniques, the latent class assignments are added to the CFA model resulting in a LISREL model with regression effects on the RIRSMACS ERS and MDRS factors. The class assignments are dichotomised and in the initial model, No RS is used as the reference category. In this model, each RSs typology is linked to each of the RIRSMACS factors included. After this, four separate models are estimated with only one class assignment (dichotomised; versus the remaining three) included as a predictor of the RIRSMACS factors. The first model allows evaluation of the joint effects whereas the other models allow evaluation of the effect of each latent class assignment separately.

The joint effects model fits adequately overall (RMSEA = 0.08, CFI = 0.98, TLI= 0.97, SRMR= 0.02), but less so with respect to the RMSEA. The effects of the latent class assignments on the RIRSMACS factors are significant except for the effect of MLRS on MDRS (p-value = 0.27). The latent class assignments jointly explain very large proportions of the variances of the RIRSMACS' ERS and MDRS factors (Table 7.4). This indicates that overall, the two models (LCA and CFA) appear to be measuring approximately the same styles. However, the sizes of the explained variances may be misleading since the RSs are correlated and each latent class assignment has an effect on each of the RIRSMACS factors.

Table 7.4: Percentages of the variances in the CFA RSs factors explained by the LCA RSs typologies

Predictor (LCA typologies)	Dependent (CFA Factors)	
	ERS	MDRS
All classes (ref: NoRS)	0.90 (0.95)	0.85 (0.92)
ERS	0.73 (0.86)	0.53 (-0.73)
MLRS	0.12 (-0.35)	0.01 (-0.10)
MDRS	0.21 (-0.46)	0.51 (0.71)
NO RS	0.01 (0.10)	0.01 (0.10)

The table shows the percentage of the variance in the CFA, RSs factors explained by the LCA RSs classes with the corresponding correlations in brackets.

Each of the four additional models with only one class as a predictor fits the data adequately with respect to most of the indices: ERS (RMSEA = 0.07, CFI = 0.99, TLI= 0.98, SRMR= 0.02), MLRS (RMSEA = 0.05, CFI = 1.00,

TLI= 0.99, SRMR= 0.01), MDRS (RMSEA = 0.07, CFI = 0.99, TLI= 0.98, SRMR= 0.02) and No RS (RMSEA= 0.10, CFI= 0.98, TLI= 0.95, SRMR = 0.03). The RMSEA value is a bit large in each of the models and it is particularly large in the model with No RS as a predictor. However, when evaluated together, the adequacy of the models is supported by a majority of the indices.

When only the ERS type (ERS= 1, all other classes= 0) is the predictor of the RIRSMACS factors, a large percentage of the variance of each factor is explained (Table 7.4). ERS compared to the other typologies, is associated positively with the ERS factor but negatively with the MDRS factor. The MLRS type explains a relatively large proportion of the variance in the ERS factor but it performs poorly as a predictor of MDRS (Table 7.4). In particular, it explains approximately 12% (correlation = -0.35) of the variance in the ERS factor and approximately 1% (correlation = -0.10) of the variance in the MDRS factor. The MDRS type predicts both MDRS and ERS in the CFA model. It explains a large proportion of the variance in the corresponding CFA factor and a relatively large proportion of the variance of the ERS factor. Finally, as should be expected, the No RS category performs very poorly as a predictor of either ERS or MDRS.

The results indicate that the LCA and CFA techniques show adequate convergent validity in the measurements of ERS and MDRS. In addition, the LCA MLRS type relates modestly to the CFA ERS factor (correlation =-0.35). This supports the position that CFA researchers need not include MLRS as a factor once ERS is included. A final note is that although No RS is explicitly included in the LCA model, it is not an RS. As such, there is no need for it to be include in the CFA correction models. In fact, it would be extremely challenging to include it as a factor at all. If researchers wish to study the absence of an RS, they may have to employ a classification technique such as LCA.

7.6.4 Respondents' Characteristics and RSs: LCA versus CFA

The effects of the respondents' characteristics (covariates) with the LCA technique (Table 7.5) are based on an MLg model whereas the results of the CFA technique (Table 7.6) are based on a LISREL model (Wang & Xiaoqian, 2012). Both the MLg model (Likelihood Ratio=1716.194, degrees of freedom=15, p-value = 0.00) and the LISREL model (RMSEA = 0.03, CFI = 0.99, TLI= 0.99, SRMR= 0.01) provide useful predictions of the RSs.

The results indicate that the proportion of the variances in the class assignments explained in the MLg model (Pseudo R-squared: Cox and Snell, 8%; McFadden, 3%) and the proportion of the variances of the RIRSMACS RSs factors (ERS, 3% and MDRS, 5%) explained by the respondents' variables are small. These small percentages are consistent with the literature on the explanatory power of the respondents' characteristics (see Van Vaerenbergh & Thomas, 2013).

Age. The results for the MLg and the LISREL model are the same for age. Age does not distinguish between any of the categories of responders (MLg model) and it does not predict any of the RIRSMACS factors (LISREL model).

Gender. In the MLg model, males are more likely than females to use No RS compared to ERS and MDRS, but gender does not distinguish among ERS, MLRS and MDRS. In the LISREL model, gender is not a predictor of ERS and MDRS. Given that both of the significant effects in the MLg model occur when No RS is involved, these results are not interpreted as real conflict between the two techniques since No RS is not included in the CFA model.

Education. In the MLg model, respondents with primary compared to secondary education are more likely to use ERS compared to MLRS and No RS and are more likely to use MDRS compared to No RS. However, education does not distinguish between ERS and MDRS and between MLRS and either MDRS or No RS. In the LISREL model, education predicts only ERS. In

Table 7.5: The effects of the respondents' variables in the LCA model

Reference Class	Predictor	Multinomial Logit Effects		
		MLRS	MDRS	No RS
ERS	Intercept	1.28**	1.96**	1.81**
	Age	−0.13	−0.10	−0.01
	Gender (ref: female)	0.34	0.19	0.54**
	More than Secondary (ref: secondary)	0.17	0.30	0.24
	Primary (ref: secondary)	−0.67**	0.33	−0.81**
	Ethnicity (ref: Minority)	0.77**	1.32**	0.46**
MLRS	Intercept		0.68	0.53
	Age		0.00	0.01
	Gender (ref: female)		−0.15	0.19
	More than Secondary (ref: secondary)		0.13	0.07
	Primary (ref: secondary)		0.34	0.13
	Ethnicity (ref: Minority)		0.55**	0.32*
MDRS	Intercept			−0.15
	Age			0.00
	Gender (ref: female)			0.34**
	More than Secondary (ref: secondary)			−0.06
	Primary (ref: secondary)			−0.47**
	Ethnicity (ref: Minority)			−0.87**

* significant at the 10% level. ** significant at the 5% level.

this case, primary education increases the use of ERS compared to secondary education (10% significance), but eduction does not affect the use of MRDS.

The results for education are consistent between the two techniques for ERS but inconsistent for MDRS. It should be noted that the 10% level of significance is required for an effect of education on ERS in the LISREL model whereas a significance level of 5% is used in most studies. Given the observed results, most researchers would therefore not report a significant effect of education from in LISREL. As a consequence, the findings from the two models would also be conflicting for ERS. The LISREL model appears to be less sensitive than the MLg model in regard to the effect of education.

Ethnicity. Ethnicity is the most predictive variable in both models. In

Table 7.6: The effects of the respondents' variables in the CFA model

Respondents' Variable	Standardised Effects	
	ERS	MDRS
Age	0.00	0.00
Gender (ref: female)	−0.10	0.08
More than Secondary (ref: secondary)	−0.05	0.07
Primary (ref: secondary)	0.15*	−0.04
Ethnicity (ref: minority)	−0.34**	0.41**

* significant at the 10% level. ** significant at the 5% level.

the MLg model, the majority ethnic group is less likely to use ERS than the other styles including No RS. Furthermore, this ethnic group is more likely to use MDRS than MLRS and No RS and more likely to use No RS than MLRS. In the LISREL model, the majority ethnic group uses less ERS but more MDRS than the minority group. These results are consistent between the two techniques.

In general, the comparisons of the LCA and CFA results for the effects of the respondents' characteristics indicate that they are very similar. A caveat to this is that a clear conflict in the results for the effect of education on ERS is encountered. Another potential issue is that the two techniques may exhibit differential sensitivity to some effects which can results in different conclusions. The application of different modelling techniques, may account for some of the observed inconsistencies in the effects of the respondents' characteristics encountered in the literature.

7.7 Discussion

Even when a confirmatory LCA (categorical) is employed, its application to modelling RSs is exploratory in the sense that the specific styles modelled are not determined before hand. ERS often emerges as either one of, or the only style in such models. ERS is also detected in this study. In contrast, ARS which is popular in CFA models is not detected. LCA also identifies two kinds of mild responding that are not popular in the literature, but which are both used by larger subgroups than ERS (see Table 7.1). Without using LCA, the importance of these RSs in the Guyanese population might not have been recognised. We would most likely have modelled ARS, ERS, DARS and MRS which are more well-known and which are often included in CFA models (RIRSMACS model) (see Weijters et al., 2008).

Using CFA with this population, Thomas et al. (2014) report low convergent validity of DARS. The current findings shed new light on this issue: DARS is not very important in the Guyanese population. Furthermore, neither ARS nor MRS appear are salient. The salient RSs are ERS, MLRS and MDRS, but CFA researchers will not be able to explicitly and meaningfully correct for MLRS and ERS simultaneously with separate factors based on indices of the kind used in the RIRSMACS model. This is due the large correlation between the two factors.

However, all is not lost. CFA researchers can exploit the large, negative correlation between MLRS and ERS. By correcting for ERS, variations due to MLRS are largely already controlled. CFA researchers therefore need only correct for ERS and MDRS in Guyanese data instead of ARS, DARS and MRS. After-all, correcting for RSs is not about maintaining the research status quo, rather, it is about reducing the errors caused by the respondents' systematic response tendencies. A caveat is that RSs demonstrate response scale specificities (Weijters, Cabooter, & Schillewaert, 2010) and as such this recommendation applies only to five-point, fully labelled scales.

The current results suggest the need for a change in the usual approach to determining the styles to model with CFA in general. The RSs modelled with CFA are determined apriori and the researcher can control the styles included. LCA can complement the representative indicators approach to modelling RSs with CFA by identifying the specific typologies that are important. This presents the opportunity for research on the salient RSs in various populations using classification techniques in order to provide the needed guidance. The stability of RSs over time (Billiet & Davidov, 2008; Weijters, Geuens, & Schillewaert, 2010) also means that researchers will not need to constantly re-discover the salient styles. There is also a need for cross-cultural research on the salient RSs typologies.

The regularity with which ERS appears in LCA models is interesting. LCA may itself have an ERS in the sense that it may be more sensitive to a tendency to use the scale endpoints. This should be investigated with simulation data. In particular, there is a need for RSs recovery studies to clarify the extend to which LCA is sensitive ERS and other RSs. Comparisons of the ability of LCA and CFA to recover the various RSs will also be useful. Such studies can assist researchers in determining which RSs, if any, are mostly likely to affect research results given the method of analysis chosen. This knowledge along with information about the salient RSs in the population will assist in determining the RSs that are most important given the population and the method of analysis.

We find high convergent validity for ERS and MDRS between LCA and CFA. The convergent validity observed is even higher than that between the representative indicators measure and the LCA, ERS style factor shown by Kieruj and Moors (2013). This is very promising for the study of RSs in general and for the use or representative indicators in particular. The measurements of RSs using the representative indicators approaches are similar between LCA and CFA as long as the salient RSs are modelled with CFA. Hence, adjustments for RSs with these approaches will be similar between the two techniques. However, to date, there are no guidelines on how to use representative indicators to correct for RSs with LCA.

Correcting for RSs with representative indicators within the LCA framework, would require a confirmatory LCA model and a set of heterogeneous items that detect RSs and which are modelled with the RIRSMACS model.

Researchers should also test their substantive model (including only content items and the content factor) to determine the number of classes present. In the third step, the RIRSMACS model should be merged with the LCA model using a factor mixture modelling approach (see Clark et al., 2013; Lubke & Muthén, 2005; B. O. Muthén, 2006, 2008, about factor mixture modelling) with the following constraints:

1. The parameters of the RIRSMACS model are constant across the latent classes.
2. The substantive LCA items load on each RS factor with equal loadings across all the items per RS factor.
3. Start the estimation with one RS factor then gradually increase the number until the best model is obtained. It is recommended that only the salient RSs be included in the RIRSMACS component of the model. These can be determined an LCA model (as done in this study).

Although the LCA and CFA models generally produce comparable results for the effects of the respondents' characteristics, one area of conflict is detected. Whereas, education has a significant effect on MDRS in the LCA model, it lacks any effect in the CFA model. In addition, the techniques show differential sensitivity to the effect of education on ERS. The technique employed (LCA or CFA) may therefore account for some of the conflicting results in the literature about the effect of the RSs antecedents. This differences in the predictions also has implications beyond the effects on RSs to any situation in which either LISREL or structural LCA models are used. This issue should be investigated further with simulated data so that researchers can understand the ways in which the two techniques differ.

The findings about the effects of the respondents' variables are generally similar to those of previous studies, but there are a few differences. While there is some support for the lack of effect of age on ERS (Moors, 2008; Thomas et al.,

2014) the results contradict the positive effect reported by Aichholzer (2013); Kieruj and Moors (2013) and Weijters, Geuens, and Schillewaert (2010). There is wider agreement with the finding that gender has no effect on ERS (Aichholzer, 2013; Kieruj & Moors, 2013; Moors, 2008; Weijters, Geuens, & Schillewaert, 2010) except that based on the latent classes, females are more likely than males to use an RS compared to No RS.

The inverse relationship between education and ERS is also consistent with some research findings (Aichholzer, 2013; Weijters, Geuens, & Schillewaert, 2010), but inconsistent with others (Kieruj & Moors, 2013; Moors, 2008). We note however, that a strictly inverse relationship is not confirmed since significant differences occur only at the lowest level of education. An interesting finding is that respondents with lower than secondary education have a general tendency to be more decisive and even when they give milder responses, they are still more likely to communicate the direction of their opinion. This is found in the LCA model.

Finally, A general higher tendency of the minority ethnicity to avoid the extreme response categories (Ayidiya & McClendon, 1990; Bachman & O'Malley, 1984; Thomas et al., 2014) is confirmed. However, the current LCA results offer further clarification. Although the majority ethnic group gives milder responses, the members of this group still tend to communicate the direction of their opinions (MDRS).

References

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel studies. *Social science research*, *42*(3), 957–70.

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Deerfield, IL: AAPOR.

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*(2), 229–247.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491–509.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research*, *36*(4), 542–562.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Bureau of Statistics. (2002). 2002 population & housing census. Retrieved from `http://www.statisticsguyana.gov.gy/census.html`

Byrne, B. (2012). *Structural equation modeling with mplus: Basic concepts, applications and programming.* New York: Taylor & Francis Group.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement in variance. *Psychological Bulletin*(3), 456–466.

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J.

(2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural equation modeling : a multidisciplinary journal*, *20*(4).

Cloud, J., & Vaughn, G. (1970). Using balanced scales to control acquiescence. *Sociometry*, *33*, 193–202.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761–775.

Fornell, C., & Larcker, D. F. (1981). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Journal of Marketing Research*, *18*, 39–50.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Hagenaars, J. A., & McCutcheon, A. (2002). *Applied latent class analysis.* Cambridge University Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Jin, K.-Y., & Wang, W.-C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement*, *74*(1), 116–138.

Kankaraš, M., Guy, M., & Vermunt, J. K. (2011). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 359–384). New York, NY: Taylor & Francis Group.

Kankaraš, M., & Moors, G. (2009). Measurement euquivalence in solidarity attitudes in europe: Insights from a multiple-group latent-class factor approach. *International Sociology*, *24*(4), 557–579.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, *40*(2), 279–310.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*(3), 320–342.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, *10*(1), 21–39.

McCutcheon, A. (1987). *Latent class analysis.* Califronia: Sage Publications Inc.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, *20*(4), 303–320.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*(6), 779–794.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101 (Suppl. 1)*, 6–16.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Evaluating structural equation models with unobservable variables and measurement error. *Structural Equation Modeling*, *14*, 535–569.

Thomas, T. D., Abts, K., & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement*, *74*(1), 97–115.

Thomas, T. D., Abts, K., & Vander Weyden, P. (in press). Measurement invariance, response styles and rural-urban measurement comparability. *Journal of Cross-Cultural Psychology*.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana.* Ghent, Belgium: Ghent University.

Wang, J., & Xiaoqian, W. (2012). *Structural equation modeling: Applications using mplus* (3rd ed.). UK: John Wiley & Sons, Ltd.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological methods*, *15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

CHAPTER 8

FACTOR MIXTURE REPRESENTATIVE INDICATORS CORRECTIONS FOR RESPONSE STYLES

Factor Mixture Representative Indicators Corrections for Response Styles in Latent Class and Factor Models

Troy Devon Thomas[1,2] and Koen Abts[3]

[1] *Department of Comparative Science of Cultures, Ghent University.*

[2] *Department of Mathematics, Physics and Statistics, University of Guyana*

[3] *Centre for Sociological Research, University of Leuven.*

Abstract

This paper presents factor mixture models (FMMs) to correct for response styles (RSs) based on representative indicators. On one hand, it demonstrates the use of an FMM to adjust for RSs in the common factor component that is estimated with categorical indicators by modelling the RSs as latent classifications. On the other hand, it demonstrates the use of an FMM to make RSs adjustments to the measurement of a categorical latent variable that is estimated in the latent class component. In this case, the RSs are modelled in the common factor component. Both approaches to modelling RSs are novel from the perspective that representative indicators adjustments for RSs have previously been restricted to traditional factor models. The models are generalisable to cases in which the substantive latent variable is estimated with a combination of ordinal and continuous outcomes.

Keywords: Response styles, representative indicators, factor mixture model, Latent class analysis, hybrid models, measurement

8.1 Introduction

Response Styles (RSs) are the respondents' systematic tendencies to respond in certain ways to rating scale items regardless of the content of the items. They bias research results and it is therefore important to correct for them whenever rating scale data are analysed (Baumgartner & Steenkamp, 2001). As factor models with categorical indicators and latent class analysis (LCA)[1] become more popular for analysing survey data, it is increasingly important that the methods of correcting for RSs are extended to both techniques taking into account the between-technique, convergent validity of the RSs measurements (Van Vaerenbergh & Thomas, 2013).

The representative indicators approach which involves the use of a separate set of heterogeneous items to measure the RSs, is promising since it shows high convergent validity between LCA and confirmatory factor analysis (CFA) (see Chapter 7). However, whereas both models for studying and correcting for RSs with representative indicators are available for CFA with continuous indicators, this approach has not been used previously to correct for RSs with either CFA with categorical manifest variables or LCA. This paper extends representative indicators corrections for RSs to both techniques using factor mixture models. It presents two examples to illustrate the methods. These approaches to adjusting for RSs serve to equip researchers with additional advanced tools to control RSs bias.

8.2 The Effects of RSs

Several RSs are used by survey respondents, but acquiescence RS (ARS: tendency to agree) and extreme RS (ERS: tendency to use the scale endpoints) are studied most often (Van Vaerenbergh & Thomas, 2013). Consequently, most of what is known about RSs are applicable to ARS and ERS. However,

[1] Here we mean latent class analysis with categorical (nominal or ordinal) latent variables.

though less well-known, disacquiescence RS (DARS: tendency to disagree) and midpoint RS (MRS: tendency to use the scale midpoint) are also among the more traditionally recognised styles.

Within the CFA (and structural equation modelling) framework, the RSs modelled are usually determined beforehand and it is common to include some combination of the four most popular RSs. This approach to determining which RSs to control ignores the cultural specificities of RSs (Harzing, 2006; Morren, Gelissen, & Vermunt, 2012a, 2012c; Thomas, Abts, & Vander Weyden, 2014). Due to cultural effects, some of the well-known styles may not be salient in all populations (Chapter 7). This is accounted for in LCA models for RSs.

LCA allows the salient RSs to emerge from the data rather than apriori decisions. The particular RSs modelled are therefore identified from the results. In this sense, the RSs component of the model is exploratory even though an overall confirmatory LCA may be employed. For example, mild RS (MLRS: tendency to avoid the scale endpoints) and mild directional RS (MDRS: tendency to avoid both the scale midpoint and endpoints) have been detected with LCA, but these are not often included in CFA models (Chapter 7; Aichholzer, 2013). In spite of potential differences in the RSs modelled with LCA and CFA, their adverse impacts on research results are confirmed with both techniques (Billiet & McClendon, 2000; Moors, 2012; Morren, Gelissen, & Vermunt, 2012b; Thomas, Abts, & Vander Weyden, in press).

RSs exist throughout the data set and they affect both the univariate and the multivariate distributions and ultimately the structure of measurement models (Billiet & McClendon, 2000). They bias factor loadings and construct means (Kankaraš & Moors, 2011; Weijters, Schillewaert, & Geuens, 2008; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). For example, higher(lower) ERS can increase(decrease) factor loadings whereas higher(lower) ARS increases(decreases) the means of manifest variables (Cheung & Rensvold, 2000). These effects are non-uniform across subgroups of respondents. For example, the mean levels of ARS, ERS, DARS and MRS differ significantly between rural and urban areas within the same country and these differentials affect within-country measurement comparability (Thomas et al., 2014, in press). RSs can hinder or result in metric and scalar invariance and can either distort or altogether conceal mean differences between groups (Morren et al., 2012b; Thomas et al., in press).

RSs also bias structural relationships. For example, Moors (2012) shows that the well-accepted gender effect on leadership styles is due to RSs. Such spurious relationships may result from deflated(inflated) variances and correlations in combination with the other effects of the RSs on the measurement models (Baumgartner & Steenkamp, 2001; Moors, 2012). Therefore, unless they are controlled, RSs are competing explanations for structural relationships between constructs (Van Vaerenbergh & Thomas, 2013).

8.3 Two Important RSs Paradigms: Nuisance and Personality

Researchers have traditionally viewed RSs as either nuisances or as meaningful personality constructs. When viewed as nuisances, the RSs are attributed to the situation and the focus is on controlling their effects while considering substantive topics. When viewed as personality constructs, the RSs are important enough to be studied, but they may also be controlled (Van Vaerenbergh & Thomas, 2013). Although, RSs have large time-invariant components (Billiet & Davidov, 2008; Weijters, Geuens, & Schillewaert, 2010), situational variables may cause some respondents to switch between styles (Aichholzer, 2013). Both of the major perspectives are therefore supported and their coexistence is not a contradiction, since situational variables may encourage or discourage the inherent tendencies of the respondents (Baumgartner & Steenkamp, 2001).

Consistent with the two major ways of viewing RSs, researchers have developed methods that either seek only to correct for their effects or methods that facilitate studying them. For example, balanced scales which include both negatively and positively worded items can correct for acquiescence RS (ARS: tendency to agree) (Baumgartner & Steenkamp, 2001; Cloud & Vaughn, 1970). Another method of controlling RSs is the use of standardisation (Fischer, 2004). A feature of correction-only approaches is that the same items measure content and style which can lead to confounding of the two. The methods developed for both studying and correcting for RSs tend to avoid such confounding of content and style by using representative indicators.

The representative indicators approaches to measuring RSs involve the use of dedicated RSs items (Greenleaf, 1992). These items do not measure a common underlying construct and they have low inter-correlations. If the respondents respond in systematic ways to such heterogeneous items, the response patterns are indicative of RSs.

Representative indicators may be obtained from a random selection of items that are already included in the questionnaire; each one from a separate construct, provided that the questionnaire includes a variety of unrelated topics (Thomas et al., in press; Weijters et al., 2008). This is possible with most survey questionnaires. However, if a researcher intends to conduct a survey on a limited set of issues that do not avail enough heterogeneous items, it would be necessary to include additional questions to facilitate RSs estimation (Van Vaerenbergh & Thomas, 2013).

8.4 Two Important Modelling Techniques: LCA and CFA

CFA and LCA are two important modelling techniques in survey research and RSs models may be implemented with both of them. The style factor approach to modelling RSs, which involves specifying an RS(s) factor(s) on the items measuring substantive constructs, may be implemented with both CFA and LCA (Billiet & McClendon, 2000; Moors, 2012). The CFA style factor (method factor) requires balanced scales and it models only ARS (Billiet & Davidov, 2008; Billiet & McClendon, 2000; Welkenhuysen-Gybels et al., 2003).

In contrast, when LCA is employed, the scales may or may not be balanced and several RSs may be modelled simultaneously (Moors, 2003, 2012; Morren et al., 2012c). The LCA, RSs factors are not determined beforehand; researchers must evaluate the model results to determine which RS(s) are detected (Moors, 2003). Researchers have successfully modelled ERS (Moors, 2003, 2004), ERS and ARS (Kieruj & Moors, 2013; Moors, 2012) and MRS (Moors, 2008) with the LCA style factor approach.

The LCA and CFA style factors cannot exist independently of substantive constructs. In this regard, the method remains true to its origins within the RSs nuisance paradigm. However, once modelled with CFA or LCA, the style factor(s) may be studied. This effectively transforms a correction method into one that both corrects for RSs and allows them to be studied. Researchers are therefore able to either correct for and study RSs using style factors when the indicators of substantive latent variables are regarded as either continuous (with CFA) or as categorical (with LCA).

In contrast, there are gaps in the literature in relation to RSs corrections with representative indicators. RSs can be modelled with representative indicators with both CFA and LCA. Within the CFA framework, the Representative Indicators Response Styles Means and Covariance Structure (RIRSMACS) model uses dedicated continuous items to model ARS, ERS, DARS and MRS and it allows for studying and controlling RSs (Thomas et al., 2014; Weijters, Geuens, & Schillewaert, 2010; Weijters et al., 2008). The model may be modified to include other RSs or to exclude some of those identified. However, there are no examples of such an approach applied to factor models with categorical indicators. Within the LCA framework, dedicated indicators have been used to study ERS, MLRS and MDRS (Chapter 7; Aichholzer, 2013), but there are no examples of RSs corrections with representative indicators when LCA is employed. These represent important limitations for researchers who simultaneously do not want to regard rating scale data for substantive latent variables as continuous and want to avoid the adverse effects of RSs (Van Vaerenbergh & Thomas, 2013). Factor mixture models avail solutions for both of these cases.

8.5 Representative Indicators Factor Mixture Models for RSs

Factor mixture models (FMMs) are hybrid models that combine LCA and common factor analysis and effectively combine the strengths of the two modelling techniques (Clark et al., 2013; B. O. Muthén, 2008). They enable accounting for population heterogeneity through the latent class component by allowing the common factor model to be estimated conditionally on the latent classes (Clark et al., 2013; Yung, 1997). FMMs may also include covariates and full structural equation models within the latent classes (Jedidi, Jagpal, & DeSarbo, 1997) and such models may be estimated on categorical or continuous outcomes (L. K. Muthén & Muthén, 1998–2012). FMMs have been successfully exploited to model social desirability and other sources of heterogeneity (Leite & Cooper, 2010; Lubke & Muthén, 2005; B. O. Muthén, 2006). In this paper, we propose the use of two FMMs: one that makes RSs corrections to the indicators of continuous latent variables and one that makes corrections to the indicators of categorical latent variables.

RSs Corrections to the Common Factor Component. To adjust the categorical indicators of substantive continuous latent variables (factors) for RSs, we propose a representative indicators factor mixture latent class response styles (RIFMLCRS) model (Figure 8.1). This model uses CFA for the substantive construct and LCA for the RSs and thereby allows both the substantive construct and the RSs to be modelled with categorical indicators. Although we emphasise adjusting for RSs when the CFA model is based on categorical indicators with the RIFMLCRS model, there is nothing inherent in the method that prevents the common factor component from containing continuous indicators.

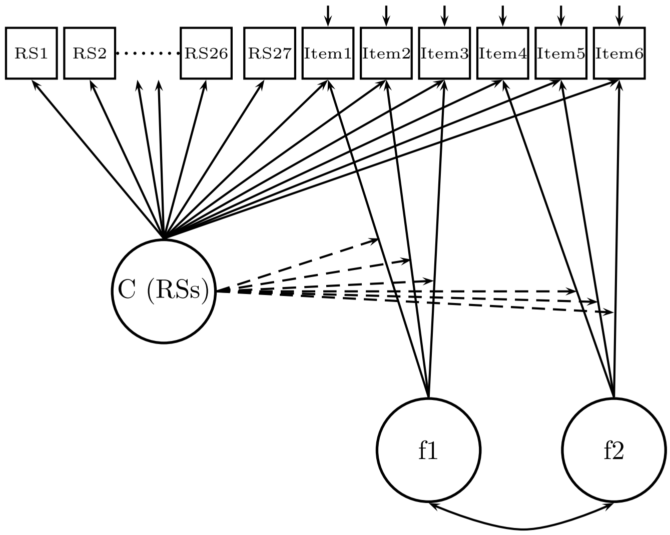

Figure 8.1: The FMRILCRS Model

To adjust for the RSs using the RIFMLCRS model, a set of heterogeneous categorical outcome variables that are shown to measure RSs when LCA is applied and a separate set of categorical (or continuous) outcome variables that measure the substantive construct under consideration are required. The two models (LCA and CFA) are then merged and estimated with the latent RSs classes determined by both the dedicated RSs indicators and the indicators of the substantive construct (see Figure 8.1).

In the estimation of the RIFMLCRS some important guidelines must be followed. The respective thresholds of the items must be equated within each latent class, but they may differ between classes in keeping with the representative indicators paradigm. In this way, the items are affected uniformly by each RS. The estimation begins with two latent classes and the number of classes is increased in a step-wise manner until the best fitting model with interpretable RSs classes is obtained. The best fitting model must have at most as many classes as identified in the separate LCA of the RSs items. At this stage, the common factor model should be freely estimated in each latent class. In particular, the latent classes of the RIFMLCRS will differentially affect the parameters of the common factor model (Dashed lines: Figure 8.1). This conditional estimation of the common factor model is in effect the adjustment for the RSs.

RSs Corrections to the Latent Class Component. When the substantive latent variable under study is categorical (labelled C) and is to be examined with LCA, the factor mixture RIRSMACS (FMRIRSMACS) model (Figure 8.2) may be used to make adjustments for RSs. The FMRIRSMACS model consists of classes determined by the indicators of the substantive, categorical latent variable. These indicators are also modelled as manifest variables of the continuous latent RSs variables which are established with the RIRSMACS model (see the methods section) (Weijters et al., 2008). Each indicator of the categorical latent variable loads on each continuous RSs variable and the loadings are equated across all the items and latent classes per RSs. In addition, all the parameters of the common factor component are constant across the latent classes. The RSs model is therefore not conditional on the latent classes. In this model, the common factor component of the FMM affects the relationships in the substantive latent class model.

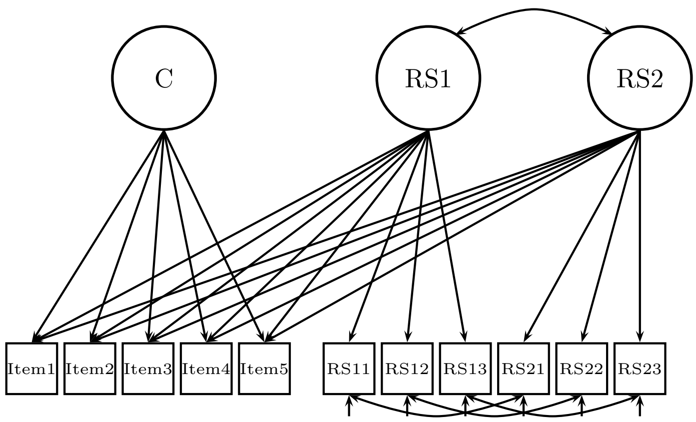

Figure 8.2: The FMRIRSMACS Model

The FMRIRSMACS model (and FMMs in general) relaxes the conditional independence assumption of LCA by allowing the RSs factors to impact on each of the indicators of the categorical latent variable (See Figure 8.2). The RSs are therefore viewed as resulting in dependence among the indicators even after the substantive categorical latent variable is modelled. Imposing conditional independence as is usually done in LCA, disregards the respondents' use of RSs. Consequently, the model parameters become biased to accommodate the RSs. This is avoided by the use of the FMRIRSMACS model.

8.5.1 Convergent Validity of RSs Measurements

An important issue to consider with different models for RSs and RSs calculated with different techniques is convergent validity. Since researchers have shown that LCA and CFA can produce different result (Kankaraš, Vermunt, & Moors, 2011), it is important to establish whether the RSs models suffer from modelling technique specificities. Modelling technique specificities in relation to RSs measurement would present difficulties in interpreting what is measured as RSs and this would ultimately overshadow the entire endeavour of measuring and correcting for RSs.

Research on the convergent validity of RSs models is limited. However, the available evidence indicates that there is low convergent validity between representative indicators measures of RSs and adhoc RSs measures. Consequently, researchers are cautioned against using adhoc measures instead of representative indicators to measure RSs (De Beuckelaer, Weijters, & Rutten, 2010).

The convergent validity between a representative indicators measure of ERS and an LCA, ERS style factor is more promising. In particular, correlations of 0.37 and 0.49 are reported (Kieruj & Moors, 2013). This suggests that measurements for ERS are potentially consistent between methods of measurement. In addition to this, there is strong convergent validity between representative indicators measures or ERS (correlation=0.86) and MDRS (correlation=0.71) between LCA and CFA (RIRSMACS) (See Chapter 7). Furthermore, the regression effects of the respondents' socio-demographic characteristics on the RSs are, with few exceptions, consistent between the two techniques (Chapter 7).

The large correlations between the LCA and CFA and the consistent predictions, especially when representative indicators are used, indicate that the two techniques tend to give the same results for RSs. This is very encouraging as it suggests that researchers can be confident in the measurement and corrections for RSs with representative indicators across CFA (RIRSMACS) and LCA. In light of this evidence, the proposed RSs corrections in the latent class and the common factor components of the FMM (RIFMLCRS and FMRIRSMACS) are similar.

8.6 Data and Methods

8.6.1 Data

The data used in this study are obtained from the Values and Poverty Study in Guyana (VAPO Guyana) which was conducted between April and May 2012. The study was funded by the Flemish Inter-University Counsel (VLIR) and jointly executed by the University of Guyana and Ghent University. It investigates both methodological and substantive issues and provides an opportunity to study RSs with representative indicators in a non-Western setting. The VAPO Guyana focused on the coastal regions (Region 2, 3, 4, 5, 6, and 10) which account for approximately 90% of the total population of the country and the data were collected via face-to-face interviews by a survey organisation (DPMC) under the supervision of the Universities of Guyana and Ghent (Vander Weyden, Abts, Thomas, Greeves, & Vereecke, 2012).

The VAPO Guyana employed a two-step sampling procedure which randomly selected municipalities with probability proportional to size, and respondents within the municipalities with equal probabilities. This procedure resulted in the selection of 87 clusters within 51 municipalities. In total, 1048 individuals were interviewed at an overall response rate of 87%. The data are weighted by iterative proportional fitting.

8.6.2 Variables

Two attitude constructs are used in the first part of the analysis. These are perceived discrimination and economic insecurity. Perceived discrimination measures feelings of fraternalistic relative deprivation emanating from perceived unequal treatment and relative shortcomings compared to other groups in regard to social resources and public policy resulting in feelings of social injustice (Abts, 2012). This construct is measured by three items which are scored on 5-point fully labelled rating scales (1 to 5: Disagree/Agree). These items are:

1. If we need something from the government, people like me have to wait longer than others.
2. People like me are being systematically neglected, whereas other groups received more than they deserve.
3. The government does a lot more for other ethnic groups than for us.

Economic insecurity refers to increased feelings of economic vulnerability and negative expectations about one's future socio-economic position (Abts, 2012). This construct is measured by three items which are scored on 5-point rating scales which have both numeric and verbal labels (1 to 5: Never, Rarely, Sometimes, Regularly, Often). The items which measure this construct are: How often are you worried that:

1. your financial worries will increase in the coming years?
2. you will have difficulties in keeping your financial position?
3. your children and the coming generation will have it much more difficult?

The measurements of perceived discrimination and economic insecurity are already validated in other surveys (see Abts, 2012; Swyngedouw, Abts, & Rink, 2009).

In the second part of the analysis, poverty attributions is used as a categorical latent variable. Work on poverty has traditionally been based on the three explanations provided by Feagin (1972). The individualistic attribution holds the individual responsible whereas the structural attribution holds external economic forces responsible and the fatalistic attribution holds poverty as being due to forces beyond the control of the individual but does not attribute it to society. These dimensions are confirmed in other studies (Feagin, 1975; Feather, 1974), but this three-tier model has been modified to include finer dimensions (Lepianka, Van Oorschot, & Gelissen, 2009; Nilson, 1981). For example, structural explanations may include both economic and non-economic factors (Furnham, 1982; Payne & Furnham, 1985) and Morçöl (1997) shows that both the individualistic and structural explanations may form two dimensions. Poverty is also viewed as due to culture (Cozzarelli, Wilkinson, & Tagler, 2001).

In this study, poverty attributions is measured by five outcome variables which tap various views that individuals may hold. Included are items that target individual elementary poverty, relative deprivation, deviant behaviour, discrimination and stratification. These are indicated by the following respective items: Poverty is a situation in which people:

1. do not have sufficient resources to provide food and clothing.
2. are not able to participate in education and health.
3. lost control over their livelihoods, and their social responsibilities towards their relatives.
4. undergo humiliation or “eye-pass”[2].
5. are confronted with the negative results of underdevelopment of the country.

[2] Guyanese colloquial term that means disrespect in this context.

These items are scored on 5-point fully labelled rating scales (1 to 5: Disagree/Agree).

8.6.3 Methods

A total of 27 5-point rating scale items (Disagree/Agree, See Appendix A.2) are used as the representative indicators for estimating the RSs. The VAPO Guyana questionnaire includes 35 RSs items which were identified following a pre-test (PAPI survey) at the University of Guyana (n=1000 students). In this pre-test, only the items with low inter-correlations ($|r| \leq 0.30$) were selected and they represent a random selection from various constructs covering several topics (including government, politics, society, crime gender roles and many more).[3] The selected items were then included in the VAPO Guyana questionnaire to ensure that separate items are always available to measure RSs in addition to the substantive constructs included (Vander Weyden et al., 2012). The 27 RSs items used have an average interitem correlation of 0.06. The scale format of these items is the same as that of the indicators of the substantive construct except economic insecurity which has different verbal labels. This mismatch of the verbal labels for this construct is a limitation of this study since that scale format may affect the RSs (Weijters, Cabooter, & Schillewaert, 2010).

RIFMLCRS Model. Before implementing the RIFMLCRS model, the continuous latent factors (perceived discrimination and economic insecurity) with categorical indicators are modelled (with CFA) to ensure that the indicators indeed measure the respective constructs. In this model the factors are correlated. The 27 dedicated RSs items are also analysed using LCA to determine which RSs they measure. In the LCA model, the respective thresholds of the 27 items are equated within each latent class.

[3]Neither pretesting nor separate items are absolutely necessary for controlling RSs. Usually, heterogeneous item can be obtained by randomly selecting one item per construct from the constructs that are included the questionnaire provided that the questionnaire covers a variety of topic areas.

Following these evaluations, the RIFMLCRS is implemented as the best fitting and interpretable mixture model consisting of the correlated continuous latent variables (CFA model) and latent RSs classes determined by both the 27 dedicated RSs items and the indicators of the continuous latent variable (LCA model). The latent classes which emerge from the data are therefore the RSs.

In the LCA component of the model, the thresholds of all the items are equated within each latent RSs class. However, the respective item thresholds are allowed to be different between the classes. In the common factor component of the model, the continuous latent variables are scaled by fixing the first factor loading to 1. Each of the remaining parameters in this component of the model is estimated conditionally on the latent classes. Therefore, essentially, a different factor model is estimated for each group determined by the RSs.

FMRIRSMACS Model. Before implementing the FMRIRSMACS model, the components are evaluated separately. The substantive categorical latent variable (types of poverty attributions) is evaluated followed by the RIRSMACS model of the salient RSs in the data.

The substantive poverty attributions model is estimated with LCA with each item loading freely on the latent variable. The number of classes are determined and if further constraints are justifiable, they are admitted. Exploratory LCA is not a requirement at this stage of the FMRIRSMACS implementation. If the model parameters are known beforehand, a confirmatory LCA maybe employed for the purpose of confirming the model.

To obtain the values of the RSs indicators of the RIRSMACS factors, the pool of 27 RSs items is divided at random into three blocks of 9 items each and one indicator per RS is calculated from each of the three blocks. In each block, the indicators are calculated as:

$$ERS = [f(1) + f(5)]/k,$$

and

$$MDRS = [f(2) + f(4)]/k$$

where $f(x)$ is the frequency of the response option x and k is the number of items per block. As a result, each RSs in the RIRSMACS model has three indicators.

The decision to model ERS and MDRS is based on the salient styles detected in the LCA of the RSs items (discussed subsequently). The RSs detected are ERS, MLRS and MDRS in addition to a group that uses no RS. However, given that MLRS and ERS are strongly correlated in the RIRSMACS model, we include only ERS and MDRS to avoid multicollinearity (See Chapter 7). Furthermore, it is also expected that if both ERS and MLRS are included, at most one of them will impact significantly on the items of the substantive construct in the RSs adjustment model.

In the RIRSMACS (CFA) model, the RSs factors correlate as are the error terms of the indicators that are calculated from the same block of items (Weijters et al., 2008). Apart form the fixing the first factor loading per factor to 1 to scale the factor and not allowing cross-loadings no other constraints are applied on the sizes of the factor loadings. However, cross-loadings are not allowed.

Following these evaluations, the FMRIRSMACS model is estimated by merging the two components into an FMM. In the FMM the outlined constraints of the FMRIRSMACS model are applied.

The models used in this paper are implemented in Mplus 7.11 with the default estimator — robust maximum likelihood — and the selection of the latent class and the FMMs is based on a combination of the AIC, BIC, adjusted BIC (ABIC) and the Mendel-Rubin Adjusted LRT in addition to the interpretability of the extracted latent classes (Kankaraš, Guy, & Vermunt, 2011; Nylund, Asparouhov, & Muthén, 2007). Interpretation is especially important in the case of the RSs classes.

8.7 Results

8.7.1 The RIFMLCRS Model

Prior to making adjustments for RSs with the RIFMLCRS model, the validity of the substantive factors under consideration and the latent RSs classifications that exist in the data need to be established. The substantive constructs examined are perceived discrimination and economic insecurity. They are modelled simultaneously and are allowed to correlate. These constructs are evaluated with a single class FMM which is equivalent to the common factor model (Clark et al., 2013). The latent RSs categories are evaluated separately with LCA.

In the factor model for perceived discrimination and economic insecurity, the factor loadings are large (Table 8.3) and the Likelihood Ratio statistic lacks significance ($L^2 = 4870.49, df = 15520$). The items are therefore valid indicators of the respective factors and the model fits the data adequately.

The RSs classifications are determined from a sequence of latent class models with an increasing number of classes. The AIC, BIC and ABIC values indicate that the overall fit of the model improves as the number of classes increases (Table 8.1). However, the Mendell-Rubin Adjusted LRT for 6 versus 5 classes lacks significance (p-value =0.13) which indicates that the inclusion of the sixth class is not necessary.

The 4-class solution (Figure 8.3) highlights three distinct RSs (ERS, MLRS and MDRS) and a class consisting of those who use no RS. The 5-class solution (Figure 8.4) adds a very weakly evidenced ARS category (labelled ARS*). In particular, the probabilities of the strongly disagree (1) and disagree categories (2) are higher than would be expected in the ARS class. Furthermore, the evidence for ARS in becomes weaker as the number of classes is increased.

The evidence for ARS remains weak in the 6-class solution and a weakly evidenced MRS category (labelled MRS*) emerges (Figure 8.5). These cat-

Table 8.1: LCA Model Selection

Classes	AIC	BIC	ABIC	Class Proportions								Entropy
				C1	C2	C3	C4	C5	C6	C7	C8	
				ERS	MLRS	MDRS	No RS	ARS*	MRS*	MDRS2	MLRS2*	
2	75330.46	75375.05	75346.46	0.19	0.81							0.96
3	73448.08	73517.44	73472.98	0.14	0.60	0.26						0.92
4	72497.39	72591.53	72531.18	0.12	0.19	0.25	0.45					0.89
5	72118.11	72237.02	72160.80	0.06	0.13	0.23	0.45	0.13				0.89
6	71909.32	72053.01	71960.09	0.06	0.12	0.19	0.45	0.13	0.06			0.89
7	71742.38	71910.84	71802.85	0.06	0.13	0.14	0.29	0.13	0.06	0.19		0.84
8	71618.70	71811.93	71688.07	0.05	0.11	0.13	0.13	0.09	0.05	0.16	0.27	0.83

Class assignment is based on the most likely class membership. MDRS: Mild Directional RS.
* indicates that the RS is not unequivocally established.

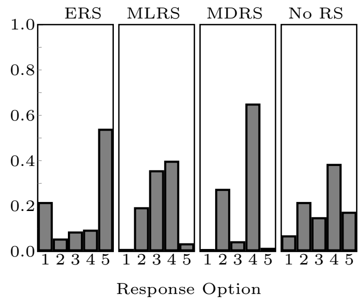

Figure 8.3: Probabilities for four-class solution

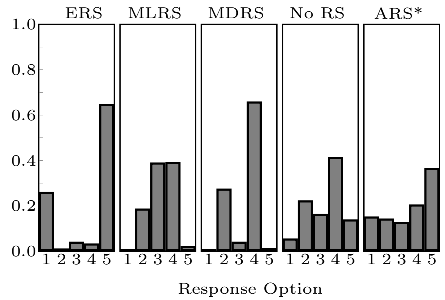

Figure 8.4: Probabilities for five-class solution

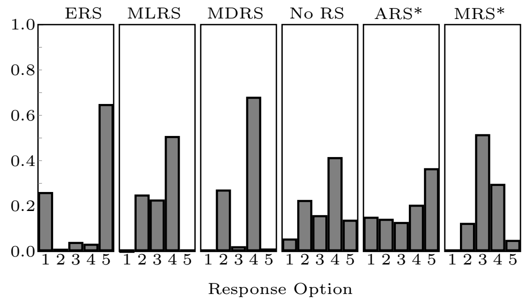

Figure 8.5: Probabilities for six-class solution

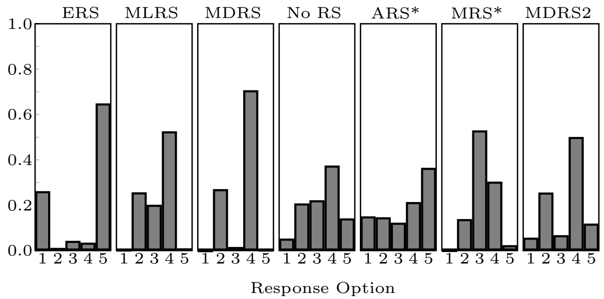

Figure 8.6: Probabilities for seven-class solution

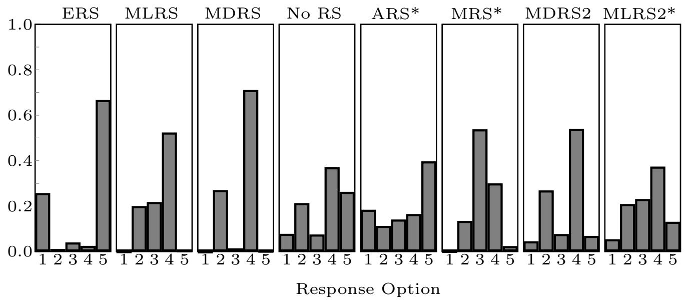

Figure 8.7: Probabilities for eight-class solution

egories are not further clarified in the 7- and 8-class solutions (see Figures 8.6 and 8.7). However, these models add further complexity. The 7- and 8-class solutions introduce refinements of classes that are already included in the model. Although this results in better fitting models, they are unnecessary given the goal of identifying and adjusting for RSs. These refinements are therefore rejected along with the 5- and 6-class solutions. Including them in the RSs correction models is likely to introduce further errors rather than correct for the RSs errors in the data.

The 4-class solution which accounts for ERS, MLRS, MDRS and no RS (Figure 8.3) is accepted as the best fitting and interpretable model. Given this selection, the RIFMLCRS model will contain at most four latent RSs classifications.

In the next step, the latent class model for the RSs and the common factor model for the substantive constructs are merged and estimated simultaneously. We begin with a 2-class model and increase the number of classes in a stepwise manner up to the 4-class model. Based on this procedure, the 3-class model is determined to be the best fit for the data (Table 8.2). We note here that the 4-class model fails to converge. However, the 3-class model still permits illustration of the use of the RIFMLCRS model. This 3-class model makes adjustments for ERS, MLRS and MDRS by computing the common factor model conditionally in each of the three latent RSs classifications of the respondents.

Table 8.2: RIFMLCRS Model Selection

Classes	AIC	BIC	ABIC	Entropy	Class Proportions		
					ERS	MLRS	MDRS
2	92769.92	92893.79	92814.38	0.95	0.20	0.80	
3	91067.70	91260.93	91137.06	0.91	0.14	0.58	0.28

Comparisons of the parameter estimates of the common factor model before and after adjusting for the RSs, highlight several differences and some

Table 8.3: Factor Loadings and Factor Variances

	Without RSs	With RSs		
		ERS	MLRS	MDRS
Perceived Discrimination (PD)				
DISC1	0.81	0.41	0.49	0.71
DISC2	0.99	0.53	0.52	0.71
DISC3	0.78	0.57	0.52	0.66
Economic Insecurity (EI)				
INSEC1	0.91	0.63	0.65	0.88
INSEC2	0.89	0.66	0.71	0.90
INSEC3	0.73	0.65	0.75	0.91
Variances				
PD	6.23	0.67	1.02	3.36
EI	15.50	2.12	2.40	11.43
Covaraince	2.78	0.80	0.48	1.44

of the differences are substantial (Table 8.3). Large difference occur for the standardised factor loadings before and after controlling the RSs. These large differences indicate that the RSs affect the loadings substantially. This is consistent with the findings of Thomas et al. (in press) based on the same data with the RIRSMACS method. We note however, that the factor loadings are in most cases inflated by the RSs.

Large difference are also observed for the factor variances and covariances before and after the RSs are included. In general, the precision of the model improves and there is less overlap between the constructs. Reduction in the covariance between constructs after controlling RSs is also reported by Moors (2012) based on the LCA style factor approach. The RIFMLCRS Model confirms that RSs affect factor loadings and the relationships among factors and it offers a way of adjusting for these effects.

In addition to the differences in the parameters before and after controlling for the RSs, there are large differences in the parameters between the latent RSs groups. The validity of the items are much higher in the MDRS group compared to the ERS and MLRS groups. In fact, the validity of two of the items measuring economic insecurity improved in the MLRS group. Overall, the items perform better in the MDRS group compared to the ERS and MLRS groups.

Between-RS differences in the factor variances and covariances are also observed. In each case, the coefficients are largest in the MDRS group. In addition to having an overall effect on the factor loadings, and variances and covariances, the RSs impact on these parameters differentially.

In view of the substantial effects of the RSs on the measurements of perceived discrimination and economic insecurity, these constructs should be studied with the RSs controlled. The RIFMLCRS offers a way of doing this and it also facilitates evaluation of how each RS affect the model parameters. It therefore appears to be a useful addition to the researcher's repertoire.

8.7.2 The FMRIRSMACS Model

The FMRIRSMACS model makes adjustments for RSs when the substantive latent variable under consideration is categorical. This model uses continuous latent RSs factors modelled with continuous outcome variables to make adjustments to the latent class component. Before establishing the FMRIRSMACS model, the substantive latent categorical variable is estimated with LCA and the RIRSMACS model for the RSs must be evaluated separately. The substantive categorical latent variable used in this section is poverty attributions.

The fit of the poverty attribution model (LCA) improves with respect to the AIC, BIC and ABIC values as the number of classes increase from 1 to 3, but the 3-class model does not provide additional information since the Mendell-Rubin Adjusted LRT lacks significance (Table 8.4). The 2-class model is therefore accepted. Apart from the difference in the sizes of the two classes, the classes differ in the intensity of the responses to the items. The first class consists of those individuals who strongly agree (5) with each of the items, while the second class consist of those who provide milder responses (mainly response option 3 to 4) (see Table 8.6). The individuals in the second class are

Table 8.4: Poverty Attributions Model Selection

Classes	AIC	BIC	ABIC	LRT	Entropy	Class Proportions		
						C1	C2	C3
1	11677.89	11776.97	11713.44					
2	9976.64	10179.74	10049.52	0.00	0.93	0.77	0.23	
3	9536.64	9843.77	9646.85	0.74	0.83	0.42	0.23	0.36
2*	10006.76	10130.60	10051.20	0.00	0.93	0.79	0.21	
2**	10336.85	10381.43	10352.84	0.00	0.92	0.79	0.21	

* Respective thresholds equated within the first class. ** Respective thresholds equated within all classes. LRT – p-value of the LRT test

still most likely to agree, but they do not agree strongly with the items.

Given the computational complexity of mixture models, it is better if this model could be simplified by equating the thresholds of the items within the latent classes. When applied to the first class, a better fitting model with respect to BIC results and the change in the adjusted BIC value is very small. However, when the thresholds are equated in the both classes, the fit of the model deteriorates (Table 8.4). The model with equal thresholds in the first class is therefore accepted and used subsequently to illustrate the implementation of the FMRIRSMACS model.

The RIRSMACS model is a CFA model which is estimated based on continuous manifest variables (Weijters et al., 2008). In this analysis, we avoid modelling the traditionally recognised RSs and focus on the salient RSs that emerge from the data: ERS, MLRS, and MDRS (see Table 8.1). However, including MLRS in the RIRSMACS component is unnecessary given that it has a large negative correlation with ERS (correlation = -0.98, Chaper 7). Furthermore, including the two will introduce multicollinearity. Therefore, only ERS and MDRS need to be modelled. The 2-factor RIRSMACS model fits adequately (RMSEA = 0.03, CFI = 1.00, TLI = 1.00, SRMR = 0.01) and the convergent validity of the factors (0.81,ERS and 0.76, MDRS) are adequate.

In implementing the FMRIRSMACS model, the 2-class structure of the substantive categorical latent variable is maintained and the RIRSMACS fac-

Table 8.5: FMRIRSMACS Model Selection

RSs	AIC	BIC	ABIC	Entropy	Class Proportions	
					C1	C2
ERS	7830.72	8004.14	7892.97	0.91	0.78	0.22
ERS and MDRS	4440.90	4683.68	4528.05	0.90	0.77	0.23

Table 8.6: Response Probabilities and Item Thresholds in the FMRIRSMACS Model

Item	Score	RSs Not Controlled				RSs Controlled			
		Class 1		Class 2		Class 1		Class 2	
		Prob	Thres	Prob	Thres	Prob	Thres	Prob	Thres
1	1	0.04	−3.27	0.00	−5.57	0.06	−2.89	0.00	−6.13
	2	0.03	−2.61	0.04	−1.97	0.06	−2.07	0.03	−3.38
	3	0.03	−2.20	0.04	−1.10	0.06	−1.60	0.04	−2.66
	4	0.13	−1.19	0.86	3.08	0.17	−0.66	0.87	2.91
	5	0.77		0.06		0.65		0.06	
2	1			0.01	−5.44			0.00	−6.62
	2			0.12	−3.45			0.11	−2.22
	3			0.08	−2.38			0.07	−1.64
	4			0.78	2.83			0.81	4.02
	5			0.02				0.02	
3	1			0.00	−5.00			0.00	−6.04
	2			0.16	−1.95			0.14	−1.85
	3			0.22	−1.39			0.20	−0.68
	4			0.59	3.88			0.63	3.71
	5			0.02				0.03	
4	1			0.00	−5.57			0.00	−6.85
	2			0.13	−1.60			0.12	−2.11
	3			0.09	−0.47			0.08	−1.48
	4			0.69	3.78			0.70	2.29
	5			0.09				0.10	
5	1			0.00	−6.08			0.00	−6.35
	2			0.14	−1.87			0.12	−2.05
	3			0.22	−1.26			0.20	−0.79
	4			0.59	2.35			0.62	2.91
	5			0.05				0.06	
Latent Mean		-1.27				−1.19			

Prob – Response probability. Thres – threshold. Note that the respective thresholds and probabilities for Class 1 are equal and are hence not repeated for each item tors are inserted to make adjustments for the RSs. The estimation begins with one RS and other styles are added in a step-wise manner until the best fitting model is obtained.

The first RS entered is ERS since it appears to be the most salient in the previous section of the analysis. With ERS included in the FMRIRSMACS model, the fit improves relative to the model without RSs (Table 8.4 and 8.5). When MLRS is controlled in addition to ERS, the fit of the model improves markedly over the previous versions of the model. The FMRIRSMACS model which adjusts for ERS and MDRS is therefore accepted as the best model.

Inclusion of the RSs results in small changes in the sizes of the latent classes (Table 8.4 and Table 8.5). In particular, the size of the class of respondents who give milder responses (Class 2) increases by 2% and the corresponding mean of class 1 is closer to that of class 2. There are also modifications of the conditional probabilities of the response options (score) and of the item thresholds (Table 8.6). For example, both the probability for the highest response category in the first class is reduced by more than 0.1. This probability is therefore less extreme. The effects on the second class appear to be less substantial than one the first class. However, controlling the RSs still results in adjustments to both the probabilities and the thresholds. In particular, the probability of assignment to the second class given that the respondent agrees (score = 4) increases after the RSs are controlled.

Given the changes observed, it is possible that the RSs can distort evaluations of measurement invariance in a multi-group analysis and may bias the factor means. These are consistent with the effects of RSs on measurement models, which have been found based on both LCA and CFA models (see Morren et al., 2012b; Thomas et al., in press). The current results therefore suggest that the FMRIRSMACS model indeed facilitates controlling the adverse effects of RSs when the substantive latent variable is to be analysed with LCA.

8.8 Discussion

The importance of correcting for RSs in the analysis of rating scale data cannot be overemphasised since RSs have several undesirable effects on measurements and substantive research outcomes (Morren et al., 2012b; Thomas et al., in press). At the same time, it is important to avoid confounding of content and style when RSs corrections are made (Van Vaerenbergh & Thomas, 2013). Although, researchers can take some steps to reduce the chances of such confounding in RSs models that use the same set of items to measure content and style, it cannot be said that the threat is eliminated in such models. For example, in applying the style factor to LCA, Moors (2012) advises that the style factors should be specified across the indicators of heterogeneous constructs. However, substantive factors included in the same model are likely to be related. As such, it is still possible that content and style are confounded.

Representative indicators for RSs offer a viable alternative, but such adjustments for RSs in either LCA or in factor models with categorical indicators have not been previously addressed in the literature. In fact, representative indicators RSs adjustments when the substantive latent constructs are based on categorical indicators in general was recently identified as an area in need of development (Van Vaerenbergh & Thomas, 2013). The two FMMs presented in this paper are therefore timely additions to the researchers' repertoire.

In the case of RSs corrections in models for continuous latent variables (common factor models), the RSs are modelled as latent classes and the substantive factor models is estimated conditionally on the RSs classes — the RIFMLCRS Model. This is essentially a multi-group factor model with the groups determined from the data. Researchers can perform all the tests that are usually done on multi-group factor models. In particular, measurement invariance of the factor models may be evaluated across the latent RSs classes to determine how the RSs influence the factor models. The factor models may also be extended into full structural equation models (Jedidi et al., 1997) with the RSs controlled. The inclusion of structural relationships also present the opportunity of evaluations of structural invariance.

Although we use the RIFMLCRS model to adjust for RSs in a CFA with categorical indicators, its relevance is not restricted to this case. The factors may have continuous or a combination of continuous and ordinal outcomes. However, it is important to maintain the same scale format across the representative RSs indicators and the indicators of the substantive constructs to control the differential effects of scale formats on the RSs (Weijters, Cabooter, & Schillewaert, 2010).

Given that FMMs facilitate exploratory factor analysis in the common factor component (McLachlan & Peel, 2000), it is possible for researchers to use the RIFMLCRS model to adjust for RSs in exploratory factor analysis models. This is an important development since it allows for RSs to be controlled in scale development. Factor loadings for instance tend to be reduced (but not always) when RSs are controlled in CFA models (Thomas et al., in press). Accounting for RSs at the scale development stage can therefore improve the validity of the constructs. The RIFMLCRS model should thus be seen as a prototype that can be generalised beyond making RSs adjustments to CFA models.

The RIFMLCRS has two important limitations of which researchers should be aware. FMMs in general require lot of computing resources and this is inherited by the RIFMLCRS as well as the FMRIRSMACS model. However, if invariant parameters are restricted between the classes, this can reduce processing time. As second issue that is specific to the RIFMLCRS model is due to the sample size. Some RSs categories may be small and if the total sample size is small, the conditional substantive model will essentially be computed on a small sample sizes. As such, the RIFMLCRS should be used on large data sets such as those available from large-scale surveys.

When the substantive latent variable is categorical and is estimated with LCA, the FMRIRSMACS model is a viable option for adjusting for the RSs. This model is a generalisation of the RIRSMACS model proposed by Weijters et al. (2008). In contrast to the RIFMLCRS model, the RSs included in the RIRSMACS component (common factor) are determined beforehand and there is the possibility that researchers will default to including the traditionally recognised RSs; namely ARS, ERS, DARS and MRS. We recommend that researchers model only those styles that are salient in the population since these salient style are expected to have the greatest impact on the data (See Chapter 7). The salient styles can be determined from a preliminary LCA of the RSs indicators before the indices for the RIRSMACS model are computed.

Although we use the FMRIRSMACS model to make correction for RSs in the latent class model that is based on categorical indicators, LCA can be done with continuous outcome variables (Wang & Xiaoqian, 2012) and the RSs corrections presented are also appropriate in this case. Furthermore, both ordinal and continuous indicators may be included in the latent class component provided that the scale format of the items matches that of the RSs indicators. Another possibility is that the FMRIRSMACS model may be use to simultaneously adjust for RSs in the LCA and the common factor components provided that substantive continuous latent variables are also included.

References

Abts, K. (2012). *Maatschappelijk onbehagen en etnopopulisme: Laatmoderniteit, burgers, ressentiment, migranten, politiek en extreemrechts* (Unpublished doctoral dissertation). Faculty of Social Sciences, University of Leuven.

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel studies. *Social science research*, *42*(3), 957–70.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research*, *36*(4), 542–562.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural equation modeling : a multidisciplinary journal*, *20*(4).

Cloud, J., & Vaughn, G. (1970). Using balanced scales to control acquiescence. *Sociometry*, *33*, 193–202.

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes toward the poor and attributions for poverty. *Journal of Social Issues*, *57*(2), 207–227.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761–775.

Feagin, J. R. (1972). Poverty: We still believe that god helps those who help themselves. *Psychology Today*, *6*, 101–129.

Feagin, J. R. (1975). *Subordinating the poor: Welfare and American beliefs.* Englewood Cliffs, NJ: Prentice Hall.

Feather, N. T. (1974). Explanations of poverty in Australian and American samples: the person, society, or fate? *Australian Journal of Psychology*, *26*, 199–216.

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, *35*(3), 263–282.

Furnham, A. (1982). The perception of poverty among adolescents. *Journal of adolescence*, *5*(2), 135–47.

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). STEMM: A general finite mixture structural equation model. *Jouranl of Classification*, *14*, 23–50.

Kankaraš, M., Guy, M., & Vermunt, J. K. (2011). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 359–384). New York, NY: Taylor & Francis Group.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social*

Sciences, *7*(2), 68–80.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, *40*(2), 279–310.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Leite, W. L., & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioral Research*, *45*(2), 271–293.

Lepianka, D., Van Oorschot, W., & Gelissen, J. (2009). Popular explanations of poverty: A critical discussion of empirical research. *Journal of Social Policy*, *38*(03), 421.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, *10*(1), 21–39.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York, NY: John Wiley & Sons.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, *20*(4), 303–320.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*(6), 779–794.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European*

Journal of Work and Organizational Psychology, *21*(2), 271–298.

Morçöl, G. (1997). Lay explanations for poverty in Turkey and their determinants. *The Journal of Social Psychology*, *137*(6), 728–738.

Morren, M., Gelissen, J., & Vermunt, J. (2012b). The impact of controlling for extreme responding on measurement equivalence in cross-cultural research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *8*(4), 159–170.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012a). Exploring the response process of culturally differing survey respondents with a response style: A sequential mixed methods study. *Field Methods*, *25*(2), 162–181.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012c). Response strategies and response styles in cross-cultural surveys. *Cross-Cultural Research*, *46*(3), 255–279.

Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101 (Suppl. 1)*, 6–16.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nilson, L. B. (1981). Reconsidering ideological lines: Beliefs about poverty in America. *The Sociological Quarterly*, *22*(2), 531–548.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Evaluating structural equation models with unobservable variables and measurement error. *Structural Equation Modeling*, *14*, 535–569.

Payne, M., & Furnham, A. (1985). Explaining the causes of poverty in the West Indies. *Journal of Economic Psychology*, *6*, 215–229.

Swyngedouw, M., Abts, K., & Rink, N. (2009). *Belgian general election study 2007. Codebook: Questions and frequency tables.* Leuven: ISPO-K.U.Leuven.

Thomas, T. D., Abts, K., & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement*, *74*(1), 97–115.

Thomas, T. D., Abts, K., & Vander Weyden, P. (in press). Measurement invariance, response styles and rural-urban measurement comparability. *Journal of Cross-Cultural Psychology*.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana.* Ghent, Belgium: Ghent University.

Wang, J., & Xiaoqian, W. (2012). *Structural equation modeling: Applications using mplus* (3rd ed.). UK: John Wiley & Sons, Ltd.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological methods*, *15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

Yung, Y. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297–330.

CHAPTER 9

CONCLUSION

Conclusion

9.1 Summary

Systematic responding to rating scale items — response styles (RSs) — bias research results (Baumgartner & Steenkamp, 2001) and unless their effects are controlled, RSs may be alternative explanations for research results that are based on rating scale data. Although rating scales are quite popular ways of obtaining opinions in surveys, RSs are generally not controlled in data analysis. This has important consequences for the accuracy of research results.

This dissertation begins with a review of the RSs literature — Chapter 3 — which highlights the known consequences of RSs and identifies several respondent and situational variables that predict and therefore offer some control of RSs. The available methods for measuring and adjusting for RSs are also discussed alongside their advantages and limitations.

Of the methods available, the Representative Indicators Response Styles Means and Covariance Structure (RIRSMACS) method (Weijters, Schillewaert, & Geuens, 2008) is found to be the most comprehensive due to its wide coverage of the RSs. This method is also flexible enough to permit other RSs to be included. However, it requires the assumption that rating scale data are continuous. For researchers who do not wish to make this assumption, the LCA style factor approach (Moors, 2003, 2012) is recommended. It should be noted however, that subsequent to these recommendations, alternatives to these methods (summarised below) were developed (Chapter 8).

In relation to the antecedents of the RSs, the respondents' sociodemographic variables inclusive of age, gender and ethnicity predict RSs depending on the RSs and the context. However, whereas such variables account for less that 10% of the variance in the RSs, culture can explain as much as approximately 75% of the RSs variance depending on the RS. Culture is therefore a major determinant of RSs.

In view of the considerable impact of culture on RSs, this dissertation investigates whether RSs across within-country subcultures should also be of major concern to researchers. If so, the relative silence of the literature on the impact of RSs in within-country research is a limitation. Two studies on RSs and subculture defined as the rural-urban divide are executed with the RSs measured by the Representative Indicators Response Styles Means and Covariance Structure (RIRSMACS) model (see Weijters et al., 2008). In both studies, the rural and urban areas in Guyana are identified based on their size, density and diversity in keeping with the Urbanism theory (Wirth, 1938).

In general, the rural-urban divide effects substantial mean differentials in acquiescence RS (ARS: tendency to agree), extreme RS (ERS: tendency to use the scale endpoints), disacquiescence RS (DARS: tendency to disagree) and midpoint RS (MRS: tendency to use the scale midpoint). Mean differences in these RSs remain after the effects of the respondents' sociodemogrphic variables are controlled and they are at least as large as the RSs differentials between data collection modes (see Weijters et al., 2008).

The RSs are also found to affect the measurements of constructs differentially between rural and urban areas. In particular, the rural-urban RSs divide can differentially bias factor loadings and item means and can either hinder or results in metric and scalar invariance. Even when measurement invariance appears to be achieved, RSs bias can still affect measurement to the extent of either distorting or altogether concealing mean differences.

The effects of the rural-urban RSs divide are similar to what can be expected in cross-cultural research (see Kankaraš & Moors, 2011) or when data are pooled across modes of collection (see Weijters et al., 2008). It is therefore as important to correct for RSs in within-country research as in cross-cultural research or in data that is pooled across collection modes. The practice of pooling within-country data across rural and urban areas without controlling RSs is therefore not justified at least in non-Western contexts.

Apart from its effects on measurement, the rural-urban divide moderates the relationships between the RSs and the respondents' sociodemographic variables. This result is an important step in understanding the possible reasons for many conflicting results in the literature. Culture moderates the relationships and it should be taken into account when interpreting such relationships.

Within-country RSs also bias structural relationships in substantive research. This is investigated with a focus on trust in institutions in Guyana and the RSs are found to bias regression relationships by either inflating the effect sizes or by resulting in entirely spurious effects (similar to Moors, 2012). This underscores the necessity of controlling RSs even in within-country research.

An issue that is highlighted by the literature review presented in Chapter 3, is the need for extensions of the representative indicators approach to correcting for RSs to latent class analysis (LCA). This dissertation addresses this issue and also extends the approach to confirmatory factor analysis (CFA) with categorical indicators. These issues are addressed in two chapters (Chapter 7 and 8).

By comparing RSs between LCA and CFA, it is determined that CFA researchers may be neglecting to control for the salient RSs by not considering the cultural context. In Guyana, ERS, mild RS (MLRS: tendency to avoid the scale endpoints) and mild directional RS (MDRS: tendency to avoid both the endpoints and midpoint of the scale) are salient. We know of no other cases in which MDRS is investigated. This suggests that LCA can be complementary to CFA since it can highlight the important styles used in the population.

Overall, the representative indicators approach to studying RSs shows high convergent validity between LCA and CFA with respect to ERS and MDRS and the effect of the respondents' sociodemographic characteristics are consistent between the two techniques. As such, researchers can be confident that the RSs modelled with the two techniques with representative indicators are similar. It is noteworthy that CFA researchers need not model MLRS once ERS is included due to the large, negative correlation between the two.

The high convergent validity of the representative indicators approach to measuring RSs between CFA and LCA makes it a good candidate for further development. In particular, although RSs may be studied using LCA (for example, see Aichholzer, 2013), examples of how to adjust for the RSs using this method are lacking.

To achieve this extension, a factor mixture model (Muthén, 2006, 2008) is employed. The RSs are implemented in the common factor component of the model using the RIRSMACS model and the substantive categorical latent variable is implemented in the LCA component. The model relaxes the conditional independence assumption of LCA by using the indicators of the substantive latent variable as indicators of the RSs as well. This is referred to as the Factor Mixture, Representative Indicators Response Styles Means and Covariance Structure (FMRIRSMACS) model.

In a second step, another model is developed with the RS in the LCA component and the substantive, continuous latent variable with categorical indicators implemented in the common factor component. The factor model is estimated conditionally on the latent RSs classes which emerge from the data. In this case, both the dedicated RSs items and the indicators of the substantive latent variable contribute to the latent classes. This model is referred to as a Factor Mixture Representative Indicators Latent Class Response Styles (FMRILCRS) model. Apart from adjusting for RSs when the substantive latent variable is modelled with CFA, the FMRILCRS model may also be applied to exploratory factor analysis and as such, it has the potential to contribute to scale development and evaluation.

Both the FMRIRSMACS and the FMRILCRS models may be extended into full structural models from which substantive research results may be obtained. The substantive items may also be regarded as begin at a combination of measurement levels provided that the number of scale points are the same as that of the RSs items.

As indicated by this summary, this dissertation moves from surveying the RSs literature to investigating the effects or RSs in within-country research, to examining the results for RSs between LCA and CFA and finally to demonstrating new methods for correcting for RSs in LCA and CFA. In the process, data from a developing country — Guyana — were collected and analysed. This contributes to advancing the agenda of conducting more data quality research in non-Western contexts in order to assist with improving data quality and what is know about it in such areas. If this dissertation achieves anything at all, we hope that it underscores the necessity of controlling for RSs in within-country research, identifies of an approach to determining the important RSs to include in CFA research and provides representative indicators approaches to adjusting for RSs in LCA and CFA models with categorical indicators.

9.2 Limitations

Although the initial intention was to base the papers in this dissertation on data from a nationwide survey of the Guyanese population, only data from the coastal regions (Region 2, 3, 4, 5, 6) and Region 10 which together account for approximately 90% of the country's population (Bureau of Statistics, 2002) were available at the time. By the completion of this dissertation, the second phase of the survey which focused on the Hinterland regions became available, but some of the articles were already published and the others close to completion. Focusing on data from the coastal regions means that an important group — Amerindians — who live mainly in the Hinterland regions are under-represented. Nevertheless, the data used are adequate for illustrating the methodological issues.

Another important limitation stems from the fact that the administrative data used for sampling was approximately 10 years old. It is therefore difficult to guarantee that the sample is representative of the population. In addition, a random walk procedure had to be used to identify the respondents. In spite of the monitoring done, the fact that the interviewers and their supervisors had to make many decisions increases the chance of bias.

A third limitation relates to the questionnaire itself. The survey was designed to test both methodological and substantive issues. As a result, some compromises had to be made between the ideal methodological requirements and the content requirements. This affected the placement of the RSs items in particular. In stead of distributing them randomly throughout the questionnaire, most of these items were placed in a battery close to the end of the questionnaire and this may have affected the responses. If RSs change drastically over the course of the questionnaire, then the analyses done in this dissertation may have overcompensated for RSs.

9.3 Recommendations for Further Research

The importance of RSs research cannot be overemphasised given the impact of RSs on research results. Although RSs have receive a lot of attention in the literature, there is still a need for much more research on this topic. In this section, recommendations that are based on the empirical research done are presented followed by recommendations that are based on the review of the literature.

There is a need for more research on within-country RSs. Such studies should investigate the rural-urban RSs divide in both Western and non-Western contexts in order to determine whether the effects encountered in this study are generalisable and further to raise awareness of the need to control RSs in within-country research. Apart from the rural-urban divide, there are other within-country variables, for example language (where applicable) that can determine subcultures, that may lead to substantial RSs differentials. These should also be investigated so that researchers understand how pooling data across such groups in the same country affects research results.

An important area for research is that of empirically reviewing existing theories with the RSs controlled. It is possible that some well-accepted relationships among variables are due to RSs (Moors, 2012). For example, this dissertation demonstrates such effects of RSs on the results of research on trust in political institutions. However, this is one study in one substantive domain in a single country using a single method of analysis. Such research is needed across a wide variety of domains, in several contexts and with different methods of analysis, for example, LCA and item response theory, so that researchers within these domains can understand how to update the current research practices.

In order to assist CFA researchers in identifying the important RSs to control in their research, it is necessary for RSs to be investigated with classification techniques. In particular, researchers should use LCA. An important area for research is that of identifying the salient RSs typologies across regions and cultures. For example, researchers can identify the salient RSs typologies across Europe and Latin America so that a large body of information is available to CFA researchers who wish to control RSs in their studies. Cross-cultural (and cross-national) comparisons of the salient RSs typologies should also be done to facilitate understanding of how the salient RSs differ across cultures (and countries). Preferably, these investigations should be done with representative indicators which avoid confounding of content and style. However, other viable methods such as the style factor (Billiet & McClendon, 2000; Moors, 2003, 2012) are also available to researchers.

Though several methods of measuring and correcting for RSs are available, there is a paucity of studies on the convergent validity of the RSs measurements across the methods. This is an area that is in need of much more research. For example, researchers can investigate the extent of convergent validity between the representative indicators approach and the style factor approaches (in LCA and CFA). The CFA style factor (method factor) should also be investigated further to establish whether it indeed measures only ARS and to determine whether it may be employed to adjust for other styles such as ERS. Researchers should use a combination of simulations and survey data in these studies since this would clarify the amount of confidence that researchers can place in these popular methods of controlling RSs.

An interesting observation is that LCA models often include ERS, but include other RSs less often. Researchers should examine whether LCA itself has an ERS in the sense that is it more sensitive to this style than other RSs. In addition, the recovery of RSs in general by LCA and CFA should be investigated with simulation data. Such studies will inform researchers about which RSs, if any, are most important to control given the method of analysis.

Two factor mixture models for controlling RSs are described in this dissertation. These models should be tested extensively to establish how useful they are for controlling RSs. One approach to this is to employ simulations, but it is also important to use survey data. Full, structural models should also be demonstrated. FMMs are quite new and many researchers may not be aware of them or may not know how to use them. Demonstrations of the applications of the models will therefore foster familiarity with and also serve to as guides to researchers on how to use them.

Researchers should demonstrate, full structural models as well as conduct comparisons of means and of measurement invariance using these FMMs. The results should also be compared with the LCA and CFA style factors in order to determine the degree of consistency between the results across the methods. All of these investigations should all be done with categorical and continuous outcome variables. The factor mixture model that makes RSs adjustments with the RSs modelled as latent classes — FMRILCRS model (see Chapter 8) — has the potential to strengthen scale development due to its applicability to exploratory factor analysis. However, the use of this model with exploratory factor analysis needs to be demonstrated.

Many conflicting results about the antecedents of RSs are encountered in the literature. Consistent with this, this dissertation finds that the effects of the respondents' characteristics sometimes vary between rural and urban areas. Furthermore, the modelling technique may also affect the results (Chapter 7). A meta-analysis that examines methodological, between-study variables and provides assessments of the different findings is necessary. Researchers should also examine the variables that mediate between the antecedents and RSs to provide insights into the cognitive processes underlying the relationships between the antecedents and RSs (Olson & Bilgen, 2011). In general, research on the antecedents of RSs has focused on investigating either stimulus-related or person-related variables (Weijters, 2006). However, Baumgartner and Steenkamp (2001) note that a person-related source of RSs (e.g., personality) may trigger or attenuate the effects of stimulus-related sources. Research should therefore examine interaction effects among antecedents.

Because we do not yet fully understand how research designs affect the use of RSs, further research on stimulus-related antecedents would be useful. Kieruj and Moors (2013) indicate that survey length might trigger ARS, but this has not yet been formally examined. Naemi, Beal, and Payne (2009) find that the amount of time a respondent spends on the questionnaire significantly influences RSs, and Cabooter (2010) investigates cognitive load (as time pressure) as a situational determinant of RSs. Other situation-related variables, such as mood, fatigue, or ego depletion, may also affect RSs, but these relationships have not been tested properly to date.

Both culture and scale format affect RSs. Merging these two issues into studies of the moderating role of culture on the effects of scale format on RSs can lead to identification of the most robust scale formats. This will be of benefit to cross-cultural research. Web surveys are becoming more popular and different colours can easily be incorporated into web-based survey instruments, but colours may impact on RSs. For example, Tourangeau, Couper, and Conrad (2007) find that for endpoint-labelled scales, when the end points are shaded in different hues compared with the same hue, responses shift toward the high end of the scales. The impact of colours on RSs should be formally examined. Research could also assess differences in RSs between unipolar and bipolar scales and between other scale formats, such as numbered and unnumbered. Tourangeau et al. (2007) indicate that the effect of shading on mean responses disappears with fully labelled scales and reduces with fully numbered scales, so there might be merit in evaluating numbered and unnumbered scales in relation to RSs. Preferably, researchers should examine all these issues in a factorial design to obtain a comprehensive picture of how scale format influences RSs.

In relation to person-related variables, researchers should further explore the effect of personality on RSs using scale-free personality tests which are not themselves contaminated with RSs. In addition, researchers should either use personality measures that do not overlap with culture (as Harzing, 2006, attempted for extraversion) or explicitly model the joint effect of personality and culture on RSs to quantify the overlap, clarify the unique effect of personality, and provide improved estimates of the explanatory power of culture for RSs.

References

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel studies. *Social science research*, *42*(3), 957–70.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Bureau of Statistics. (2002). 2002 population & housing census. Retrieved from `http://www.statisticsguyana.gov.gy/census.html`

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures* (Unpublished doctoral dissertation). Ghent University.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European*

Journal of Work and Organizational Psychology, *21*(2), 271–298.

Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101 (Suppl. 1)*, 6–16.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of personality*, *77*(1), 261–86.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, *75*, 99–114.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*(1), 91–112.

Weijters, B. (2006). *Response styles in consumer research* (Unpublished doctoral dissertation). Ghent University.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, *44*(1), 1–24.

APPENDIX

A.1 RSs Items used in Chapter 4

All items are scored in the following scale: 1 = Completely Disagree; 2 = Disagree; 3 = Neither Agree nor Disagree; 4 = Agree; 5 = Completely Agree

Striving for personal success is more important than providing for good relations with your fellowman.

I approve of people participating in legal demonstrations.

In my daily life, I seldom have time to do the things I really enjoy.

Doctors keep the whole truth from their patients.

Citizens should spend at least some of their free time helping others.

Nowadays businesses are only interested in making profits and not in improving service or quality for customers.

Men should take as much responsibility as women for the home and children.

I am satisfied with the way democracy works in Guyana.

When there are children in the home, parents should stay together even if they don't get along.

I never seem to have enough time to get everything done in my job.

I am a quiet and shy person.

Torturing a prisoner in a Guyanese prison is never justified, even if it might provide information that could prevent a terrorist attack.

When jobs are scarce, men should have more right to a job than women.

Schools must teach children to obey authority.

Employees often pretend they are sick in order to stay at home.

On the whole, my life is close to how I would like it to be.

If I help someone, I expect some help in return.

There are people in my life who really care about me.

If you want to make money, you can't always act honestly.

The prison breaks reflect the failure of the judicial system.

For crimes such as murder and drug traffic, young people from 14 years onwards should be sentenced just as adults.

Economic growth always harms the environment.

Participation of citizens in issues concerning the society should be enhanced.

Guyana is suffering from an economic crisis.

I trust the media in Guyana.

Generally, I am in good health.

Modern science can be relied on to solve our environmental problems.

The standard of living of pensioners in Guyana is acceptable.

The tax authorities are efficient at things like handling queries on time, avoiding mistakes and preventing fraud.

The Guyanese government, more than the private sector, should be primarily responsible for creating jobs.

The level of crime that we have now represents a threat to our future wellbeing.

People like me are being systematically neglected, whereas other groups received more than they deserve.

I feel myself powerless and at the mercy of current changes.

These days, you really don't know who you can trust.

Nowadays, politics has a total lack of common sense.

Same-sex couples should have the right to marry.

All politicians are profiteers.

The parliament does not succeed in solving problems, it is therefore better to abolish it.

The people should govern directly rather than through elected representatives.

The differences between classes ought to be smaller than they are at the present.

Poverty is a situation in which people are confronted with the negative results of underdevelopment of the country.

Poverty can only be solved by more equality in international relationships between countries.

A.2 RSs Items used in Chapters 5 – 9

All items are scored on the scale: 1 Completely Disagree, 2 Disagree, 3 Neither Agree nor Disagree, 4 Agree, 5 Completely Agree

I approve of people participating in legal demonstrations.

In my daily life, I seldom have time to do the things I really enjoy.

Doctors keep the whole truth from their patients.

Citizens should spend at least some of their free time helping others.

Nowadays businesses are only interested in making profits and not in improving service or quality for customers.

Men should take as much responsibility as women for the home and children.

I am satisfied with the way democracy works in Guyana.

When there are children in the home, parents should stay together even if they don't get along.

I am a quiet and shy person.

Torturing a prisoner in a Guyanese prison is never justified, even if it might provide information that could prevent a terrorist attack.

When jobs are scarce, men should have more right to a job than women.

Schools must teach children to obey authority.

Employees often pretend they are sick in order to stay at home.

On the whole, my life is close to how I would like it to be.

If I help someone, I expect some help in return.

There are people in my life who really care about me.

If you want to make money, you can't always act honestly.

For crimes such as murder and drug traffic, young people from 14 years onwards should be sentenced just as adults.

Economic growth always harms the environment.

Participation of citizens in issues concerning the society should be enhanced.

Guyana is suffering from an economic crisis.

I trust the media in Guyana.

Generally, I am in good health.

The standard of living of pensioners in Guyana is acceptable.

The tax authorities are efficient at things like handling queries on time, avoiding mistakes and preventing fraud.

The Guyanese government, more than the private sector, should be primarily responsible for creating jobs.

The level of crime that we have now represents a threat to our future well-being.

A.3 Variable List: Values and Poverty Study in Guyana

ADMINISTRATIVE VARIABLES

Item	Label	Description
EntryID	EntryID	ID of the data entry personnel
Region	Region	Region
Areatype	Areatype	Area type: rural, urban or suburban
B2_NDC	NDC	Neighbourhood Democratic Council (Municipality)
B2_Village	Village	Village code
B2_Enum	Enum	Enumerator (interviewer) number
B2_Resp	Resp	Respondent number (linked to interviewer number)
B4	roof	Material used for roof
B5	gend	Gender of respondent
Weeg_edu		Weight variable (agecateg*gender, education)
Weeg_vote		Weight variable (agecateg*gender,voting)
Timer 1		

SOCIO-DEMOGRAPHIC VARIABLES

Item	Label	Description
Q1	yrbrn	Year of birth
Q2	guyanese	Have Guyanese Nationality
Q3	nationality	Other Nationality
Q4	wrkabroad	Intentions of working abroad
Q5	maritalstatus	Marital Status
Q6	dependu18	Number of dependents younger than 18 in household
Q7	depen18to65	Number of dependents aged 18-65 in household
Q8	dependover65	Number of dependents older than 65 in household
Q9	ageatfirstchild	Age when first child was born

Q10_1	ethnicity	Ethnicity
Q10_2	ethnicityfather	Father's ethnicity
Q10_3	ethnicitymother	Mother's ethnicity
Q10_4	ethnicitypartner	Partner's ethnicity
Q11	religious	Religious
Q12	demomination	Religious denomination
Q13	religiousservice	Church attendance
Q14	familyabroad	Has family living abroad
Q15	supportoverseas	Support from overseas relatives
Q16	supportoverseas2	Type of overseas support
Q17	supportlocal	Support from local persons and organisations
Q18_1	supportgovernment	Support from Guyanese government
Q18_2	supportorg	Support from religious organisation, charity or NGO
Q18_3	supprotfamily	Support from family
Q18_4	supportneighbours	Support from neighbours
Q18_5	supportfriends	Support from friends
Q19	supporthouselot	House lot from government
Q20	hlvled	Highest level of education
Q21	agelftsch	School leaving age
Q22_1	hlvledfather	Father's highest level of education
Q22_2	hlvledmother	Mother's highest level of education
Q22_3	hlvledpartner	Partner's highest level of education
Q23	agefirstjob	Age at first job
Q24	employactive	Employment status
Q25	employsituation	Employment situation
Q26	employcomp	Type of organisation/employment
Q27	job	Job description
Q28	hhincome	Total household income per month

Q29	ownhouse	Owner
Q30_1	havcomputer	Have a computer
Q30_2	havwashmachine	Have a washing machine
Q30_3	havrefrigerator	Have a refrigerator
Q30_4	havgenerator	Have a backup electricity generator
Q30_5	havbathtub	Have bath tub
Q30_6	havflush-toilet	Have a flush toilet
Q30_7	havvehicle	Have a vehicle
Q30_8	havoutboard	Have an outboard motor
Q31_1	eatrice	Eat rice at least four times a week
Q31_2	eatcassava	Eat cassava at least four times a week
Q31_3	eatgprovision	Eat provision at least four times a week
Q31_4	eatwwbread	Eat bread at least four times a week
Q31_5	eatfarine	Eat farine at least four times a week
Q31_6	eatflour	Eat flour at least four times a week
Q31_7	eatpeas	Eat peas at least four times a week
Q31_8	eatpotatoes	Eat potatoes at least four times a week
Q31_9	eatpasta	Eat pasta at least four times a week
Timer 2		
POLITICS AND SOCIETY		
Q32_1		Meetings of religious organisation
Q32_2		Meetings of community group
Q32_3		Meetings of political parties or political organisation
Q33		Level of safety
Q34_1		because of skin colour
Q34_2		because of accent
Q34_3		because of economic situation
Q34_4		because of gender

Q35_1	Wait longer than others
Q35_2	Systematically neglected
Q35_3	Government does more for other ethnic groups
Q36_1	Can't do anything about most things
Q36_2	Future in own hands
Q36_3	Don't have a lot of control over society
Q36_4	Feel powerless
Q37_1	Financial worries will increase
Q37_2	Difficulties keeping financial position
Q37_3	Coming generation will have it more difficult
Q38_1	Don't know who to trust
Q38_2	Can't be too careful in dealings
Q39_1	Brotherhood and solidarity are nonsense
Q39_2	Personal success more important that good relations
Q39_3	Better to take care first and only for oneself
Q40_1	To solve problems, get rid or immoral, crooked people
Q40_2	Obedience to authority and respect are most important
Q40_3	Tighten laws; too much freedom is not good
Q41_1	Mixture of races is good (or bad)
Q41_2	Immigrants take jobs away
Q41_3	Immigrants undermine cultural life
Q41_4	Immigrants worsen crime problems
Q41_5	Immigrants put strain on economic system
Q41_6	The proportion of immigrant will become greater threat

Q41_7	Better is immigrants maintain their distinct customs
Q42_1	Run for public office
Q42_2	Right to marry
Q43_1	Justice system
Q43_2	Guyana defence force
Q43_3	Parliament
Q43_4	National Government
Q43_5	Guyana Police Force
Q43_6	Mass Media
Q43_7	National Elections
Q43_8	Political Parties
Q43_9	Actual President
Q43_10	Mayor's office/NDC chairman
Q43_11	Regional Democratic Counsel
Q44	Level of interest
Q45_1	Follow political parties in media
Q45_2	Discuss politics
Q46	Job performance o parliamentarians
Q47_1	Voting makes no sense
Q47_2	Parties only interested in vote; not my opinion
Q47_3	Politicians only promise a lot
Q47_4	Politicians are profiteers
Q47_5	Most politicians are competent
Q47_6	Politics lacks common sense
Q47_7	Power needs to be returned to the people
Q47_8	Need strong leader who does what the majority thinks
Q48	President of the US

Q49	Regions of Guyana
Q50	Term of Government
Q51	President of Guyana
Q52	Registered to vote
Q53	Voted
Q54	Party voted for
Q55	Party identified with
Q56	Offered favour
Q57	Favour affected vote
Q58	Satisfaction
Q59	How democratic
Q60_1	Better to abolish parliament; does not solve problems
Q60_2	Parties create more problems than they solve
Q60_3	Democracy is the best system
Q60_4	Need strong leader; not have to bother with parliament & elections
Q60_5	Democracy can exist without parliament
Q61	Democratic or authoritarian
Q62_1	President: Limit the voice of the opposition to progress
Q62_2	President: Govern without parliament if it obstructs
Q62_3	President: Ignore supreme court if it obstructs
Q62_4	People: Govern directly; no representatives
Q62_5	Those who disagree with the majority represent a threat
Q63	Electoral democracy is best
Q64_1	Corruption widespread; In politics

Q64_2	Corruption widespread; Police officers
Q64_3	Corruption widespread; Government employees
Timer 3	
SOCIAL INEQUALITY	
Q65	Expected future economic situation
Q66	Satisfaction with current, household economic situation
Q67	Country's economic situation
Q68	Choose: freedom or equality
Q69_1	Trade unions have to be more aggressive
Q69_2	Workers have to struggle for equal position in society
Q69_3	Class differences ought to be smaller
Q69_4	Difference in high and low income should remain
Q70_1	Incomes should be more equal
Q70_2	Reducing income differences; government's responsibility
Q70_3	Government should provide decent standard of living
Q70_4	Government should spend less on benefits
Q71	Just-unjust: People with higher incomes can buy better health care
Q72	Just-unjust: People with higher incomes can buy better education
Q73_1	Government or the people should provide for themselves
Q73_2	Competition is good
Q73_3	Hard work (or luck and connections) brings better life

Q74	Type of society
Q75	Ideal type of society
Timer 4	
POLITICAL CHOICES	
Q76	Left-right political leanings
Q77	Liberal-conservative leanings
Q78_1	High unemployment
Q78_2	Corruption
Q78_3	A lot of crime
Q79	Justified or not: when country facing difficult times
POVERTY	
Q80_1	Insufficient resources for food and clothing
Q80_2	Unable to participate in education and health
Q80_3	Lost control over livelihood and social responsibility
Q80_4	Undergo humiliation and eyepass
Q80_5	Faced with negative effects of underdevelopment
Q81	Choose definition of poverty
Q82_1	Drink too much or do drugs
Q82_2	Lazy and lack willpower
Q82_3	Not motivated
Q82_4	Lack intelligence and talent
Q82_5	Victims of stigmatisation and discrimination
Q82_6	Do not earn enough
Q82_7	Exploited
Q82_8	Discontinue education too soon
Q82_9	Do not have a voice
Q82_10	Don't get the same chances as others

Q83_1	Drink too much and use drugs
Q83_2	A way of life
Q83_3	Punishment from God
Q83_4	Breakdown of family and community life
Q83_5	Inadequate familial support
Q83_6	Failure of educational system
Q83_7	Individual bad luck or disability
Q83_8	Insufficient employment levels
Q83_9	Low wages
Q83_10	Inadequate social benefits
Q83_11	Government inefficiency and incompetence
Q83_12	Inequality in society
Q83_13	Unavoidable part of modern life
Q84_1	There will always be poverty
Q84_2	Increase in social welfare and pension benefits
Q84_3	More equality in international relationships
Q84_4	Increased quality of education
Q84_5	The poor don't deserve help
Q84_6	Increased job opportunities
Q84_7	Increased taxes for the rich
Q84_8	Higher minimum wage
Q84_9	Developing the interior
Q84_10	More investment from international donors
Q85	Gave financial assistance in the past six months
Q86_1	Gave to relatives
Q86_2	Gave to Family-friends
Q86_3	Gave to someone in neighbourhood
Q86_4	Gave to a stranger
Timer 5	

RSs BATTERY

Q87_1	Approve of participation in legal demonstration
Q87_2	Seldom have time to for things I enjoy
Q87_3	Doctors keep the whole truth
Q87_4	Citizens should spend free time helping others
Q87_5	Businesses just interested in profits; not service/quality improvement
Q87_6	Men should take responsibility for children as women do
Q87_7	Satisfied with how democracy works
Q87_8	Parents should stay together for the children
Q87_9	Not enough time for things in my job
Q87_10	Quiet and shy person
Q87_11	Torturing prisoners is never justified
Q87_12	When jobs are scarce, men should have more right
Q87_13	Schools much teach obedience
Q87_14	Employees feign illness to stay at home
Q87_15	Life close to ideal
Q87_16	Expectation of help reciprocation
Q87_17	People care about me
Q87_18	To make money you can't always act honestly
Q87_19	Prison breaks: failure of judicial system
Q87_20	Young people prosecuted as adults for crimes like murder and trafficking
Q87_21	Economic growth harms environment
Q87_22	Citizen participation should be enhanced
Q87_23	Guyana is suffering from economic crisis
Q87_24	Trust the media in Guyana
Q87_25	Generally in good health

Q87_26		Modern science will solve environmental problems
Q87_27		Pensioners have acceptable living standards
Q87_28		Tax authorities efficient at handling queries and preventing fraud
Q87_29		Government more than private sector has job creation responsibility
Q87_30		Crime level: threat to future wellbeing
RESPONDENT'S EVALUATION OF SURVEY		
Q88		Pleasant
Q89		Interesting
Q90		Future participation
Timer 6		
CONSTRUCTED VARIABLES		
AGE	age	The age of the respondent in years
AGEcateg		Categorised version of age variable (to compare with available census data)
EDU2	Education	Recoded highest level of education (Q20: hlvled) variable
Vote_ipf		Recoded voting (Q54) variable

References

Abts, K. (2012). *Maatschappelijk onbehagen en etnopopulisme: Laatmoderniteit, burgers, ressentiment, migranten, politiek en extreemrechts* (Unpublished doctoral dissertation). Faculty of Social Sciences, University of Leuven.

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel studies. *Social science research*, *42* (3), 957–70.

Albaum, G., Roster, C., H, Y. J., & Rogers, R. D. (2007). Simple rating scale formats: Exploring extreme response. *International Journal of Market Research*, *46*, 633–650.

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Deerfield, IL: AAPOR.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, *66* (3), 374–392.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40* (6), 1235–1245.

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54* (2), 229–247.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48* (2), 491–509.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European journal of public health*, *20* (5), 543–548.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Bean, C. (2003). Citizen confidence in social and political institutions in a changing world. In C. Bailey, D. Cabrera, & L. Buys (Eds.), *Social change in the 21st century conference.* Brisbane: Centre for Social Change Research, Queensland University of Technology.

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: a two-dimensional interpretation of acquiescence. *Psychological bulletin*, *76*(3), 186–204.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research*, *36*(4), 542–562.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(4), 608–628.

Blanco, L. R. (2013). The impact of crime on trust in institutions in Mexico. *European Journal of Political Economy*, *32*, 38–55.

Blanco, L. R., & Ruiz, I. (2013). The impact of crime and insecurity on trust in democracy and institutions. *American Economic Review*, *103*(3), 284–288.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, *71*(5), 814–833.

Bouckaert, G., & Van De Walle, S. (2001). Government performance and trust in government. Paper presented at the European Group for Public

Administration Annual Conference, 5–8 September, Vaasa, Finland.

Brengelmann, J. C. (1960). Extreme response set, drive level and abnormality in questionnaire rigidity. *The British Journal of Psychiatry*, *106*(442), 171–186.

Bulmer, M. (2001). General introduction. In M. Bulmer & D. P. Warwick (Eds.), *Socal research in developing countries: Surveys and censuses in the third world* (pp. 3–24). London: Taylor & Francis Group.

Bureau of Statistics. (2002). 2002 population & housing census. Retrieved from `http://www.statisticsguyana.gov.gy/census.html`

Byrne, B. (2012). *Structural equation modeling with mplus: Basic concepts, applications and programming.* New York: Taylor & Francis Group.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement in variance. *Psychological Bulletin*(3), 456–466.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155–175.

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures* (Unpublished doctoral dissertation). Ghent University.

Chang, E. C. C., & Chu, Y.-h. (2006). Corruption and trust: Exceptionalism in Asian democracies? *The Journal of Politics*, *68*(2), 259–271.

Chen, C., Lee, S.-y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, *6*(3), 170–175.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? the impact of making inappropriate comparisons in cross-cultural research.

Journal of personality and social psychology, *95*(5), 1005–1018.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*(1), 1–27.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Christensen, T. O. M., & Læ greid, P. (2005). Trust in government: The relative importance of service satisfaction, political factors, and demography. *Public Performance & Management Review*, *28*(4), 487–511.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*, 64–73.

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural equation modeling : a multidisciplinary journal*, *20*(4).

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality*, *15*(1), 137–152.

Cleary, P. D., Mechanic, D., & Weiss, N. (1981). The effecit of interviewer characteristics on responses to a mental health interview. *Journal of Health and Social Behavior*, *22*(2), 183–193.

Cloud, J., & Vaughn, G. (1970). Using balanced scales to control acquiescence. *Sociometry*, *33*, 193–202.

Couch, a., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of abnormal and social psychology*, *60*(2), 151–74.

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes toward the poor and attributions for poverty. *Journal of Social Issues*, *57*(2), 207–227.

Crandall, J. E. (1973). Sex differences in extreme response style: Differences in frequency of use of extreme positive and negative ratings. *The Journal of Social Psychology*, *89*(2), 281–293.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761–775.

De Jong, M. D., Steenkamp, J.-B. E., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149.

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, *16*(1), 20–30.

Feagin, J. R. (1972). Poverty: We still believe that god helps those who help themselves. *Psychology Today*, *6*, 101–129.

Feagin, J. R. (1975). *Subordinating the poor: Welfare and American beliefs.* Englewood Cliffs, NJ: Prentice Hall.

Feather, N. T. (1974). Explanations of poverty in Australian and American samples: the person, society, or fate? *Australian Journal of Psychology*, *26*, 199–216.

Fischer, C. S. (1975). Toward a subcultural theory of urbanism. *American Journal of Sociology*, *80*(6), 1319–1341.

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, *35*(3), 263–282.

Fornell, C., & Larcker, D. F. (1981). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Journal of Marketing Research*, *18*, 39–50.

Furnham, A. (1982). The perception of poverty among adolescents. *Journal of adolescence*, *5*(2), 135–47.

Gans, H. (1962). Urbanism and suburbanism as ways of life: A re-evaluation of definitions. In A. M. Rose (Ed.), *Human behavior and social processes; an interactionist approach* (pp. 625–648). Boston: Houghton-Miffl.

Gibbons, J. L., Zellner, J. a., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by spanish-english bilinguals. *Cross-Cultural Research*, *33*(4), 369–381.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Grimm, S. D., & Church, A. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, *33*(4), 415–441.

Hagenaars, J. A., & McCutcheon, A. (2002). *Applied latent class analysis.* Cambridge University Press.

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192–203.

Hamm, J. A., PytlikZillig, L. M., Tomkins, A. J., Herian, M. N., Bornstein,

B. H., & Neeley, E. M. (2011). Exploring separable components of institutional confidence. *Behavioral sciences & the Law*, *29*(1), 95–115.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, *21*(1), 111–121.

Hofstede, G. (2001). *Cultures consequences: Comparing values, behaviors, institutions and organizations across nations* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research*, *38*(1), 52–88.

Horn, J. L., & McCardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144.

House, R. J., Hanges, P. W., Javidan, S. A., Dorfman, M., & Gupta, V. (2004). *Culture, leadership, and organizations. the globe study of 62 societies.* Thousand Oaks, CA: Sage.

Hox, J. J., de Leeuw, E. D., & Kreft, I. G. G. (1991). The effect of interviewer and resondent characteristics on the quality of survey data. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 439–461). New York: Wiley.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Huang, K.-p., Lee, F.-y., & Lin, T.-m. (2013). Partisanship and institutional trust a comparative analysis of emerging democracies in East Asia. *Taiwan Journal of Democracy*, *9*(1), 47–71.

Hug, S., & Spörri, F. (2011). Referendums, trust, and tax evasion. *European Journal of Political Economy*, *27*(1), 120–131.

Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology*, *132*(2), 201–210.

Hutchison, M. L., & Johnson, K. (2011). Capacity to trust? institutional capacity, conflict, and political trust in Africa, 2000-2005. *Journal of Peace Research*, *48*(6), 737–752.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). STEMM: A general finite mixture structural equation model. *Jouranl of Classification*, *14*, 23–50.

Jin, K.-Y., & Wang, W.-C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement*, *74*(1), 116–138.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*(2), 264–277.

Jordan, L. A., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, *44*(2), 210–222.

Kankaraš, M., Guy, M., & Vermunt, J. K. (2011). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 359–384). New York, NY: Taylor & Francis Group.

Kankaraš, M., & Moors, G. (2009). Measurement euquivalence in solidarity attitudes in europe: Insights from a multiple-group latent-class factor approach. *International Sociology*, *24*(4), 557–579.

Kankaraš, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija*, *43*(2), 121–136.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across europe. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *7*(2), 68–80.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, *40*(2), 279–310.

Khodyakov, D. (2007). Trust as a process: A three-dimensional approach. *Sociology*, *41*(1), 115–132.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*(3), 320–342.

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*(1), 193–211.

Kiesler, S., & Sproull, L. S. (1986). Response effects in electronic surveys. *Public Opinion Quarterly*, *50*, 402–413.

Knowles, E. S., & Condon, C. a. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379–386.

Krosnick, J. a. (1999). Survey research. *Annual review of psychology*, *50*, 537–567.

Leite, W. L., & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioral Research*, *45*(2), 271–293.

Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research*, *23*(3), 361–

373.

Lepianka, D., Van Oorschot, W., & Gelissen, J. (2009). Popular explanations of poverty: A critical discussion of empirical research. *Journal of Social Policy*, *38*(03), 421.

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of personality and social psychology*, *2*(6), 907–9.

Lipps, O., & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods*, *23*(2), 156–172.

Listhaug, O. (1984). Confidence in institutions: Findings from the Norwegian Values Study. *Acta Sociologica*, *27*(2), 111–122.

Locke, K. D., & Baik, K.-D. (2009). Does an acquiescent response style explain why Koreans are less consistent than Americans? *Journal of Cross-Cultural Psychology*, *40*(2), 319–323.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, *10*(1), 21–39.

Lühiste, K. (2006). Explaining trust in political institutions: Some illustrations from the Baltic states. *Communist and Post-Communist Studies*, *39*(4), 475–496.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, *23*(4), 498–509.

Matarazzo, J. D., & Herman, D. O. (1984). Relationship of education and IQ in the WAIS-R standardization sample. *Journal of Consulting and Clinical Psychology*, *52*(4), 631–634.

Maxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research. In N. Schwarz & S. Sudman (Eds.), *Context effects in social psychology* *research* (pp. 279–296). New York: Springer.

McCutcheon, A. (1987). *Latent class analysis.* Califronia: Sage Publications Inc.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York, NY: John Wiley & Sons.

McQuitty, S. (1997, January). Effects of employing ridge regression in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4* (3), 244–252.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44* (7), 1539–1550.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58* (4), 525–543.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, *57* (2), 289–311.

Meuleman, B., Davidov, E., & Billiet, J. (2009). Changing attitudes toward immigration in Europe, 20022007: A dynamic group conflict theory approach. *Social Science Research*, *38* (2), 352–365.

Millsap, R. E. (1995). Measurement invariance predictive invariance and the duality paradox. *Multivariate Behavioral Research*, *30*, 577–605.

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single factor case. *Psychological methods*, *2*, 248–260.

Millsap, R. E. (2007). Invarince in measurement and prediction revisited. *Psychometrika*, *72*, 461–473.

Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4* (1), 5–9.

Millsap, R. E., & Yun-tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479–515.

Mishler, W., & Rose, R. (1997). Trust , distrust and skepticism: Popular evaluations of civil and political institutions societies in post-communist societies. *The Journal of Politics*, *59*(2), 418–451.

Mishler, W., & Rose, R. (2001). What are the origins of political trust?: Testing institutional and cultural theories in post-communist societies. *Comparative Political Studies*, *34*(1), 30–62.

Mishler, W., & Rose, R. (2005). What are the political consequences of trust?: A test of cultural and institutional theories in Russia. *Comparative Political Studies*, *38*(9), 1050–1078.

Moors, G. (2003). Diagnosing response style behavior by means of a latent class factor approach socio-demographic correlates of gender role attitudes and peceptions. *Quality & Quantity*, *37*(3), 277–302.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, *20*(4), 303–320.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*(6), 779–794.

Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, *22*(1), 93–119.

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*(2), 271–298.

Morçöl, G. (1997). Lay explanations for poverty in Turkey and their determinants. *The Journal of Social Psychology*, *137*(6), 728–738.

Morren, M., Gelissen, J., & Vermunt, J. (2012b). The impact of controlling for extreme responding on measurement equivalence in cross-cultural research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *8*(4), 159–170.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012a). Exploring the response process of culturally differing survey respondents with a response style: A sequential mixed methods study. *Field Methods*, *25*(2), 162–181.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012c). Response strategies and response styles in cross-cultural surveys. *Cross-Cultural Research*, *46*(3), 255–279.

Mõttus, R., Allik, J., Realo, A., Rossier, J., Zecca, G., Ah-Kion, J., ... Johnson, W. (2012). The effect of response style on self-reported conscientiousness across 20 countries. *Personality & social psychology bulletin*, *38*(11), 1423–36.

Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101 (Suppl. 1)*, 6–16.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of personality*, *77*(1), 261–86.

Neale, M. C., Lubke, G., Aggen, S. H., & Dolan, C. V. (2005). Problems with using sum scores for estimating variance components: contamination and measurement noninvariance. *Twin research and human genetics: The official journal of the International Society for Twin Studies*, *8*(6), 553–68.

Nilson, L. B. (1981). Reconsidering ideological lines: Beliefs about poverty in America. *The Sociological Quarterly*, *22*(2), 531–548.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Evaluating structural equation models with unobservable variables and measurement error. *Structural Equation Modeling*, *14*, 535–569.

Oberski, D. (2008). Jrule for Mplus. Retrieved from `https://github.com/daob/JruleMplus/wiki`

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, *75*, 99–114.

Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Formal definitions of measuement bias and explanation bias clarify measurement and conceptual perspectives. *Journal of Clinical Epidemiology*, *62*, 1126–1137.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Payne, M., & Furnham, A. (1985). Explaining the causes of poverty in the West Indies. *Journal of Economic Psychology*, *6*, 215–229.

Peil, M. (2001). Situational variables. In M. Bulmer & D. P. Warwick (Eds.), *Socal research in developing countries: Surveys and censuses in the third world* (pp. 71–88). London: Taylor & Francis Group.

Petković, J. (2007). Traditional values and modernization challenges in forming urban and rural culture. *Philosophy, Sociology and Psychology*, *6*, 23–39.

Poznyak, D., Meuleman, B., Abts, K., & Bishop, G. F. (2013). Trust in american government: Longitudinal measurement equivalence in the ANES, 19642008. *Social Indicators Research*.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*(2), 230–243.

Rohrschneider, R., & Schmitt-Beck, R. (2002). Trust in democratic institutions in Germany: Theory and evidence ten years after unification. *German Politics*, *11*(3), 35–58.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*(3), 129–156.

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, *25*(2), 189–197.

Rothstein, B., & Stolle, D. (2008). The state and social capital: An institutional theory of generalized trust. *Comparative Politics*, *40*(4), 441–459.

Rowland, M. L., Naidoo, S., Abdulkadir, R., Moraru, R., Haung, B., & Pau, A. (2010). Perceptions of intimidation and bullying in dental schools: A multi-national study. *International Dental Journal*, *60*, 106–112.

Saris, W. E., Satorra, A., & Coenders, G. (2004a). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311–347.

Saris, W. E., Satorra, A., & Coenders, G. (2004b). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311–347.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation*

Modeling: A Multidisciplinary Journal, *16*(4), 561–582.

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *94*, 347–363.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.

Schwartz, N., Oyserman, D., & Peytcheva, E. (2010). Cognition, communication and culture: Implications for the survey response process. In J. A. Harkness et al. (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (pp. 177–190). New Jersey: Wiley & Sons.

Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 143–179). Boston, MA: McGraw-Hill.

Secor, A. J., & Loughlin, J. O. (2005). Social and political trust in Istanbul and Moscow: A comparative analysis of individual and neighbourhood effects. *Transactions of the Institute of British Geographers*, *30*(1), 66–82.

Shevlin, M., Miles, J., & Bunting, B. (1997). Summated rating scales. a monte carlo investigation of the effects of reliability and collinearity in regression models. *Personality and Individual Differences*, *23*(4), 665–676.

Shlapentokh, V. (2006). Trust in public institutions in Russia: The lowest in the world. *Communist and Post-Communist Studies*, *39*(2), 153–174.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*(1), 50–61.

Smith, T. W. (2011). Refining the total error perspective. *International Journal of Public Opinion Research*, *23*, 464–484.

Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, *43*(4), 599–616.

Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *The Journal of Social Psychology*, *122*(2), 151–156.

Stukovsky, R., Palat, M., & Seldlakova, A. (1982). Scoring position styles in the elderly. *Studia Psychologica*, *24*, 145–154.

Swyngedouw, M., Abts, K., & Rink, N. (2009). *Belgian general election study 2007. Codebook: Questions and frequency tables.* Leuven: ISPO-K.U.Leuven.

Thomas, T. D., Abts, K., & Vander Weyden, P. (2014). Response styles and the rural–urban divide. *Educational and Psychological Measurement*, *74*(1), 97–115.

Thomas, T. D., Abts, K., & Vander Weyden, P. (in press). Measurement invariance, response styles and rural-urban measurement comparability. *Journal of Cross-Cultural Psychology*.

Tittle, C. R. (1989). Influences on urbanism: A test of predictions from three perspectives. *Social Problems*, *36*(3), 270–288.

Tittle, C. R., & Grasmick, H. G. (2001). Urbanity: Influences of urbanness, structure, and culture. *Social Science Research*, *30*(2), 313–335.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*(1), 91–112.

Usunier, J.-C. (2011). Language as a resource to assess cross-cultural equivalence in quantitative management research. *Journal of World Business*, *46*(3), 314–319.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Van der Veld, W. M. (2008). Judging misspecifications in structural equation models using JRule. *Symposium on Testing Structural Equation Models at the ISA-RC33, 7th International Conference on Social Science Methodology. Naples (Italy), September 1–5*, 1–5.

Van der Veld, W. M., & Saris, W. E. (2011a). Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Methods and applications in cross-cultural analysis* (pp. 117–148). New York: Taylor & Francis Group.

Van der Veld, W. M., & Saris, W. E. (2011b). Cause of generalized social trust. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 207–247). New York: Routledge: Taylor & Francis Group.

Vander Weyden, P., Abts, K., Thomas, T., Greeves, O., & Vereecke, J. (2012). *Codebook: Values and poverty study in guyana*. Ghent, Belgium: Ghent University.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications Inc.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*, 29–37.

Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, *54*(2), 119–135.

Van Rosmalen, J., van Herk, H., & Groenen, P. J. (2010). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research*, *47*(1), 157–172.

Wang, J., & Xiaoqian, W. (2012). *Structural equation modeling: Applications using mplus* (3rd ed.). UK: John Wiley & Sons, Ltd.

Watkins, D., & Cheung, S. (1995). Culture, gender, and response buis: An analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, *26*(5), 490–504.

Weijters, B. (2006). *Response styles in consumer research* (Unpublished doctoral dissertation). Ghent University.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Weijters, B., Geuens, M., & Schillewaert, N. (2009). The proximity effect: The role of inter-item distance on reverse-item bias. *International Journal of Research in Marketing*, *26*(1), 2–12.

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*(2), 105–121.

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological methods*, *15*(1), 96–110.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing*

Science, *36*(3), 409–422.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Lkert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*(6), 956–972.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, *44*(1), 1–24.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse worded items confound measures in cross cultural consumer research? the case of the material values scale. *Journal of Consumer Research*, *30*(1), 72–91.

Yung, Y. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297–330.

Zuckerman, M., & Norton, J. (1961). Response set and content factors in the california F Scale and the parental attitude research instrument. *Journal of Social Psychology*, *2*, 199–210.